Ensemble Postprocessor (EnsPost)
User's Manual

Version: OHD-CORE-16.2.1
Release Date: 7 April 2017

National Weather Service
Office of Hydrologic Development

Contents

1   Overview

The HEFS Ensemble Post-processor (EnsPost) applies a statistical model in order to adjust streamflow ensembles to account for hydrologic model bias and hydrologic uncertainty. It requires parameters that must be estimated prior to executing the model via the EnsPost Parameter Estimator (EnsPostPE) and are stored in a central directory for access by the model when executed. The model is executed via the HEFSEnsPostModelAdapter, which is configured as any other module within a CHPS workflow.

Read the *HEFS Overview and Getting Started Manual* for an introduction to HEFS as a whole and EnsPost in particular.

1.1   EnsPost Software Components

The HEFS ensemble post-processor (EnsPost) software consists of the parameter estimator, the EnsPostPE, and the ensemble post-processor CHPS model adapter, the EnsPost. The EnsPostPE estimates the parameter and calculates the error statistics, whereas the EnsPost post-processes the ensemble members of the model forecast over the forecast horizon. Figure 1 shows the schematic of the EnsPost components (under an assumption of a 24-hour calibration time step).
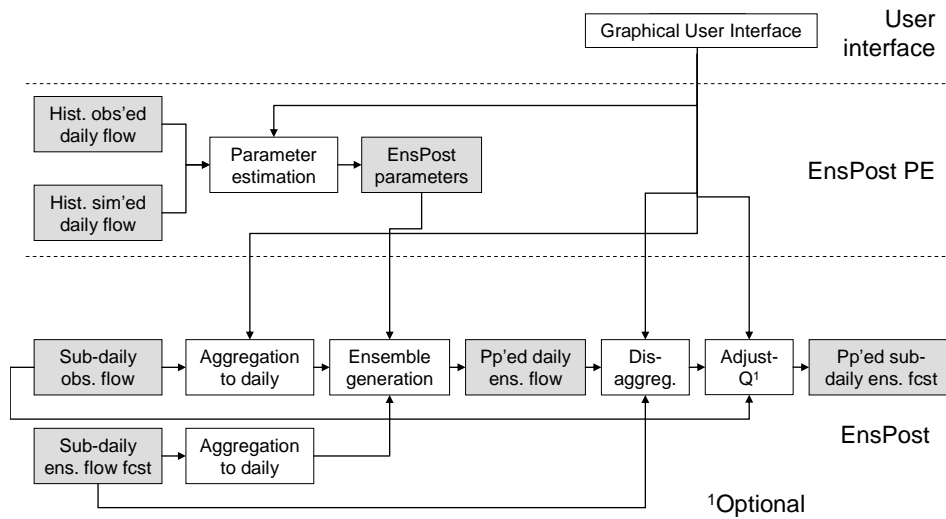

Figure 1: Schematic of the EnsPost components

1.2   Manual Layout

This document follows this outline:

Section 2: Guidance for parameter estimation through the EnsPostPE and ensemble generation through the EnsPost.

Section 3: A description of the science underlying EnsPost.

Section 4: A reference for all of the components and the bells and whistles of the EnsPostPE.

Section 5: A description of how to execute the EnsPost operationally to post-process stream flow ensembles and configuration reference manuals model adapters.

1.3 Terminology

The following terminology is used throughout this manual.

- *calibration time step*: The time step at which parameters are estimated, or calibrated.
- *parameter estimation stand-alone (SA)*: The standalone in which the EnsPostPE component was be installed; see the *EnsPostPE Configuration Guide*.
- *CHPS locationId*: The locationId used in the CHPS configuration files to specify a location.
- *CHPS parameterId*: The parameterId used in the CHPS configuration files to specify a data type. Common parameterIds referred to are as follows:
  - o MAP/FMAT: observed/forecast 6h accumulated precipitation
  - o MAT/FMAT: observed/forecast 6h instantaneous temperature
  - o TFMX/TMAX: observed 24h maximum temperature
  - o TFMN/TMIN: observed 24h minimum temperature
- *EnsPost location*: A location for which the EnsPost is to be executed and parameter estimated. An EnsPost location is defined by a CHPS locationId and parameterId and will sometimes be referred to by its *identifier* within this manual, which is "<locationId> (<parameterId>)". For example "CNNN6DEL (SQIN)".

1.4 Notation

The following notation is used:

- Important terms are displayed in *italics* the first time they are used and defined.
- Graphics user interface components are displayed in **Bold**.
- List items, such as available plug-ins or allowed parameter settings, will be in "quotes".
- Column names in tables will be in 'single quotes'.
- Text which is to be entered at a command line or into an ASCII text file (including XML files) is denoted in this font.
- Parameter names are displayed as normal text.
- XML elements are displayed in this font.
- Directory and file names will be denoted in this font.

1.5 Directories of Note

The following directories will be referred to in this manual:

- *<PE SA region_dir>*: The *parameter estimation stand-alone* (see the *MEFPPE Installation Guide*) region home directory, typically "##rfc_sa".

- *<PE SA configuration_dir>* The Config directory under *<region_dir>* (i.e., *<region_dir>*/Config).
- *<ens_post_root_dir>*: The directory selected to hold the EnsPost .tgz parameter files; see the *EnsPostPE Configuration Guide*.
- *<enspostpe_run_area>*: The directory in which the EnsPostPE will store run time information. Typically <region_dir>/Models/hefs/hefsEnsPostPERunArea. See the *EnsPostPE Configuration Guide*.

2   Guidance

The EnsPost generates a post-processed ensemble streamflow forecast using parameters estimated via the EnsPostPE. This section provides general guidance for the use of EnsPostPE to estimate parameters and the use of the EnsPost CHPS adapter, the HEFSEnsPostModelAdapter, to generate post-processed ensemble forecasts. This guidance aims to assist in making decisions about the application of the EnsPostPE and the EnsPost. It does <u>not</u> provide instructions for how to configure or execute either piece of software, although some configuration guidance will be provided (from a high level). For details about the configuration of EnsPostPE, see the *EnsPostPE Configuration Guide*. For details about the configuration of the EnsPost, see the *EnsPost Configuration Guide*.   Details about how to execute the EnsPostPE and the HEFSEnsPostModelAdapter are provided in Sections 4 and 5.1 of this manual, respectively.

2.1   Parameter Estimation with EnsPostPE

Parameters are estimated for the EnsPost with the EnsPostPE. The configuration of the EnsPostPE is described in the *EnsPostPE Configuration Guide*. Historical observations and simulations used to estimate parameters are acquired from the parameter estimation SA through the FEWS PI-service (see Section 4.2.4.3 of this manual). Options for parameter estimation are specified through the **Estimation Options Subpanel** of the EnsPostPE (see Section 4.6.2). Parameters can be estimated for a single segment or many segments simultaneously, and diagnostic tools are provided for examining the input data, as well as the resulting parameters (see Sections 4.6.3and 4.9).

This section provides guidance pertaining to the data requirements of the EnsPostPE and estimation options.

2.1.1   Data Requirements

Summary of Recommendation: At least 20 years of historical observations and simulations should be used to estimate the parameters of the EnsPost, providing the climatology of the simulated and observed flows is broadly representative of the current (and likely future) streamflow conditions.

Experience has shown that reliable estimates of the EnsPost parameters typically require at least 10 years of historical data (daily) and, ideally, 20 years or more. If the cross-correlation between the observed and simulated streamflow is particularly strong (e.g., >0.9), these requirements may be relaxed somewhat. However, if the auto-correlation or "persistence" of the observed streamflow is also strong (typical of snow basins), additional data may be required, because many samples will contain "shared" information. Also, when estimating parameters for the EnsPost with more than two seasons (e.g., a monthly calibration) or with a flow threshold that differs from the median value (i.e. 50% in each partition), additional data may be required for a reliable calibration, particularly for "tail" events (i.e., extreme low and high flows).

With all statistical models, including the EnsPost, a trade-off exists between the amount and consistency of the data used to estimate the model parameters. When calibrating the EnsPost, a long and consistent record is preferred. However, if the streamflow regime has changed "substantially" over the historical period (e.g., due to new or changing river regulations), the calibration data should be curtailed,

otherwise the EnsPost will be "correcting" for biases and hydrologic errors that are no longer applicable. Managing this trade-off requires experience of the local hydrologic conditions and datasets, for which few generalizations can be made. However, as with all decisions about the configuration and calibration of the HEFS, it should be informed by hindcasting and validation.

Displaying the historical time-series of hydrologic simulations and corresponding observations in the EnsPostPE may aid in evaluating the consistency of the hydrologic modeling errors and biases over time. However, the EnsPost diagnostics are currently very limited, an issue identified in CHPS Redmine Issue 24609. See Section 4.5.1.2 for a description of the time series diagnostics currently available.

See Section 4.5.2.1for an explanation of how to use the **Select Series Dialog** of the **Export Historical Data Subpanel** to restrict the years of data acquired when exporting historical simulation and observations time series via the FEWS PI-service for use in EnsPostPE.

2.1.2 Historical Simulation

Summary of Recommendation: Select a retrospective (historical) hydrologic simulation that it is consistent with the operational hydrologic forecast. Also, the temporal scale of the observations, simulations, and forecasts must be equivalent at the timestep used to calibrate the EnsPost and to correct the operational forecast; for example, an instantaneous flow cannot be compared with an average flow.

The historical simulations obtained through the FEWS PI-service for parameter estimation must be consistent with the streamflow forecasts that are corrected by the EnsPost, operationally (see Section 2.2 also). In other words, the historical simulations should encompass the same sources of hydrologic uncertainty and bias as the operational streamflow forecasts. For example, if streamflow observations are routed downstream, operationally, the EnsPost should be calibrated with historical simulations that are generated with the same routing procedure. Any discrepancies between the historical simulations and the operational forecasts, such as those introduced by runtime modifications, (miss)handling of river regulations, or the incorrect use of blending/ADJUST-Q, could lead to adjustments by the EnsPost that are sub-optimal, at best, and may *reduce* the quality of the raw streamflow forecasts, at worst.

When calibrating the EnsPost, and applying the EnsPost parameters operationally, the temporal scale of the hydrologic simulations, observations and (adjusted) forecasts must be equivalent. The EnsPostPE provides only limited options to aggregate input datasets and no options to disaggregate or interpolate the inputs. Thus, for example, it is not possible to calibrate the EnsPost with instantaneous hydrologic simulations at a 6h timestep unless the streamflow observations are also available at the same times, and as instantaneous quantities. However, it is possible to calibrate the EnsPost at a daily aggregation if one or both inputs can be aggregated, arithmetically, to corresponding quantities. For example, daily observed streamflows (QME) may be provided together with hourly or 6-hourly hydrologic simulations (QINE). In that case, the EnsPostPE will aggregate the QINE by forming an arithmetic mean of the QINE values that fall within the observed period (using a five-point average). More complex aggregations, disaggregations, or interpolations must be conducted externally (e.g. within CHPS), before the EnsPost is calibrated.

Identify the appropriate simulations and change the PI-service configuration file, as needed. See Section 4.2.1 for a description of the inputs to EnsPostPE including the historical simulation; Section 4.2.4.3 for an explanation of the FEWS PI-service connection; and Section 2.4.1 of the *EnsPostPE Configuration Guide* for a description of the file to modify once the historical simulation has been selected.

2.1.3 Seasons

Summary of Recommendation: Examine the seasonality of the historical observations and estimated parameters for a small, representative sample of segments for which the EnsPost will be applied and select the seasons accordingly. Apply to other segments as appropriate.

The main considerations in choosing a seasonal stratification are the characteristics of the streamflow modeling errors and the sample size available to estimate them. The objective of seasonal stratification is to separate out the modeling errors or "scatter" (i.e. the scatter between the simulated and observed flows) so that the resulting groups appear to be internally homogeneous, while being different from the other groups. For example, if there are no seasonal variations in the hydrologic modeling errors, no seasonal stratification is required (i.e. a single, "annual season" should be used). If more than one season is selected, then there must be sufficient data to estimate the EnsPost parameters. Hence, as more seasons are added, the record of historical observations and simulations must be sufficiently large to properly characterize each season. When defining multiple seasons, a historical record of 20 years or more should be considered, although hindcasting and validation will allow for a more objective assessment at a particular location.

Diagnostics available in the EnsPostPE for examining seasonality require, first, estimating the parameters of the EnsPost for "monthly seasons" (i.e., 12 seasons, each corresponding to a single month). After doing so, the parameters must be loaded and the **Diagnostic Display Panel** used to examine the estimated parameters. See Sections 4.6.3 and 4.9 for more information on the **Diagnostic Display Panel** as it is used for estimated parameters.

For example, this is a plot of the mean and coefficient of variation for the segment WGCM8 at NWRFC:

Parameter Type	Month #	Minimum	Maximum
QIN Observed Monthly Mean	12	21.313	296.084
QIN Observed Monthly Coefficient of Variation	12	0.425	1.144
QIN Simulated Monthly Mean	12	20.677	301.591
QIN Simulated Monthly Coefficient of Variation	12	0.377	0.84

Diagnostics Panel

Location: WGCM8 (SQIN)
Forecast Source: PI-xml Data
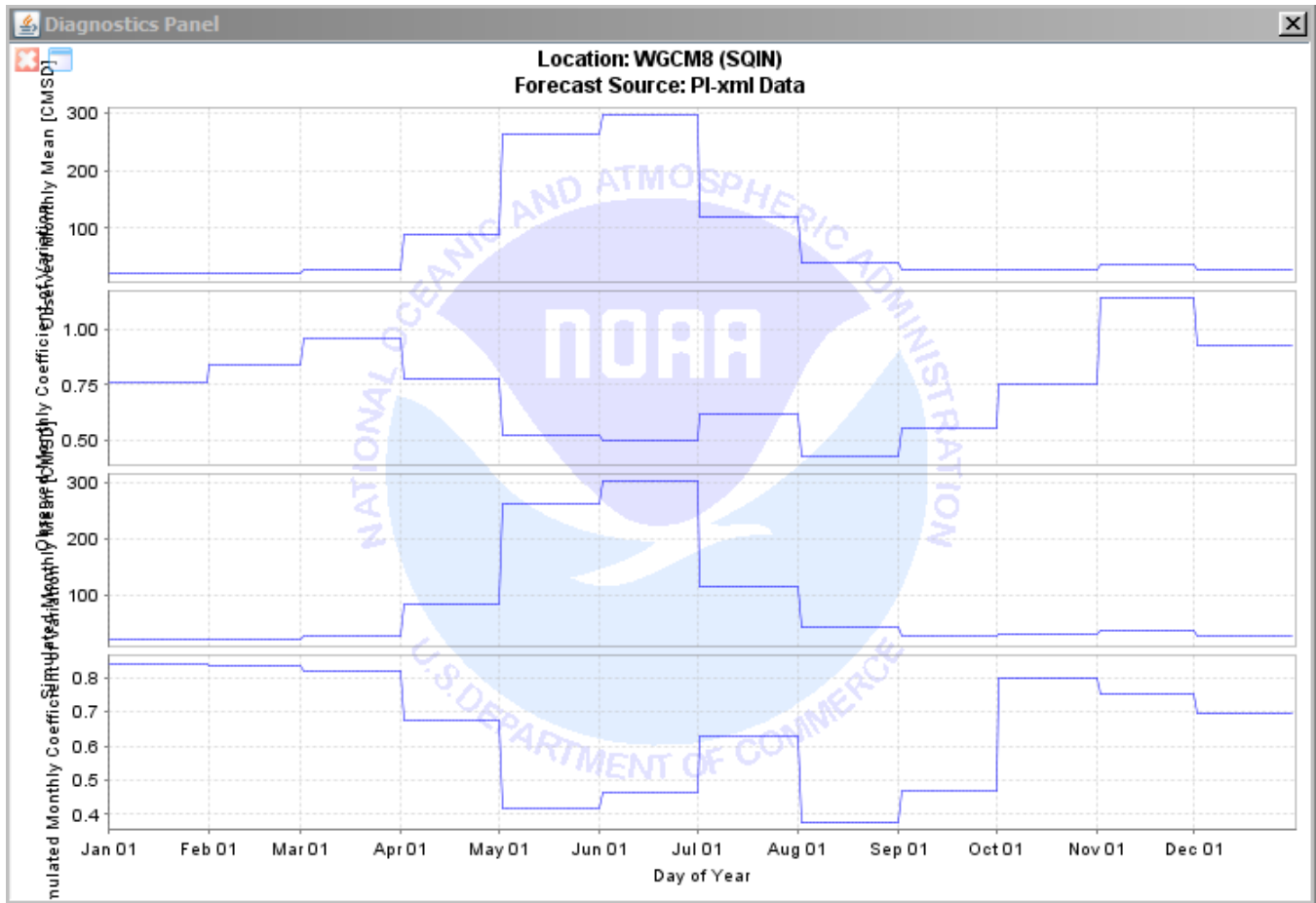
In this case, the segment is greatly impacted by snow melt in the late Spring and early Summer months.

This is a plot of the corresponding values of the *b-coefficient* for above ("Hi") and below ("Lo") the threshold (discussed in the next section). The *b-coefficient* is the coefficient of the error model that multiplies the model simulation (the *a-coefficient* multiplies the previous observed value and is equal to $1.0 - b$, so only one coefficient needs to be examined):

QIN Linear Model Coefficient A Lo	12	0.742	0.980
QIN Linear Model Coefficient B Lo	12	0.014	0.258
QIN Error Term E1 Lo	12	-0.061	0.075
QIN Error Term E2 Lo	12	0.066	0.253
QIN Linear Model Coefficient A Hi	12	0.027	0.942
QIN Linear Model Coefficient B Hi	12	0.058	0.973

10

Location: WGCM8 (SQIN)
Forecast Source: PI-xml Data
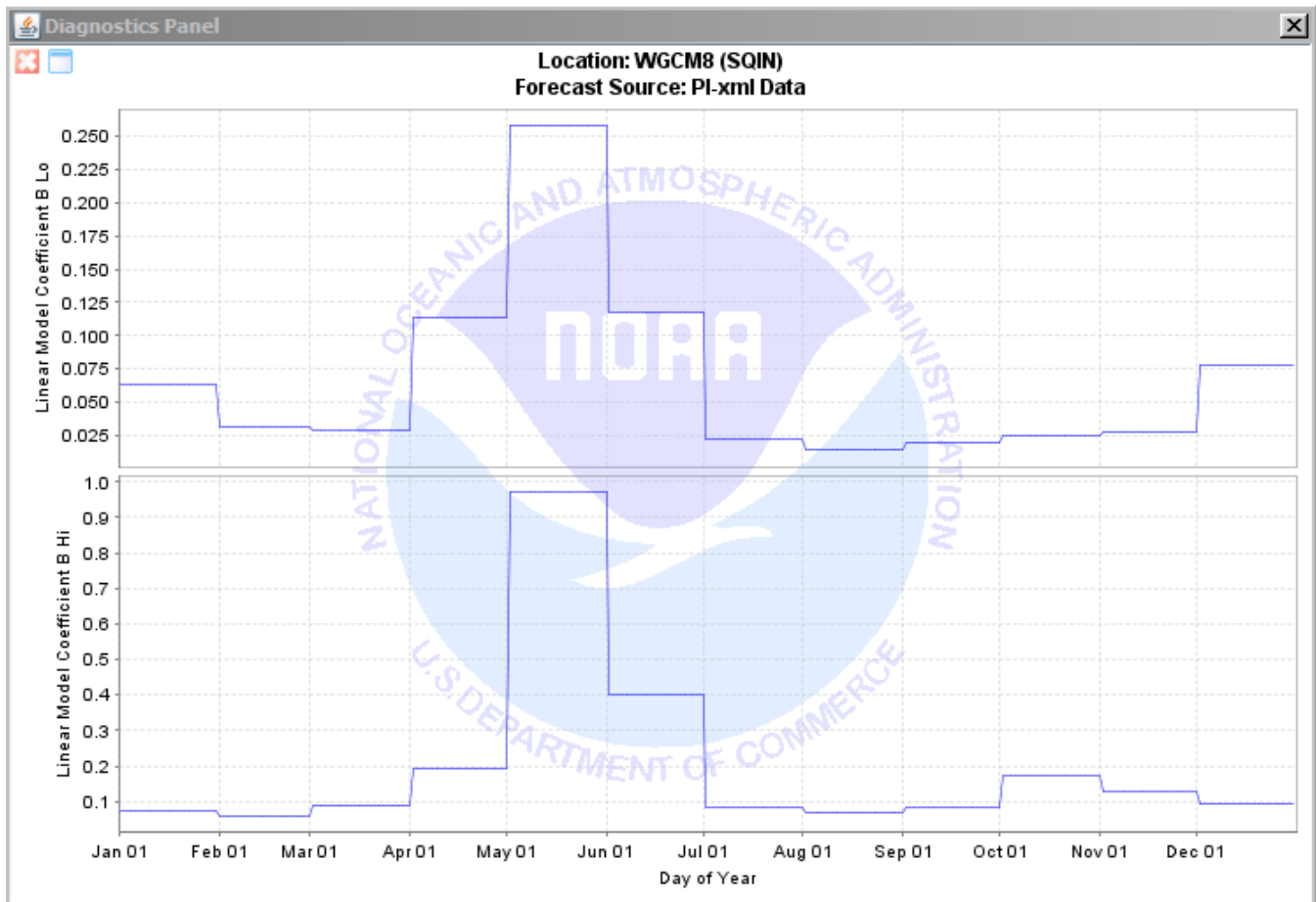
In general, the *b-coefficient* is low for months not impacted by snowmelt, implying strong persistence in the error model, and high for months impacted by snowmelt.

This is a plot of the values of the *e1* parameter (mean of the residual error in the ARX(1) model employed in the error model) and *e2* parameter (variance of the residual error) for both above and below the threshold:

| QIN Error Term E1 Lo	12	-0.061	0.075
QIN Error Term E2 Lo	12	0.066	0.253
QIN Linear Model Coefficient A Hi	12	0.027	0.942
QIN Linear Model Coefficient B Hi	12	0.058	0.973
QIN Error Term E1 Hi	12	-0.086	0.001
QIN Error Term E2 Hi	12	0.111	0.504

Location: WGCM8 (SQIN)
Forecast Source: PI-xml Data
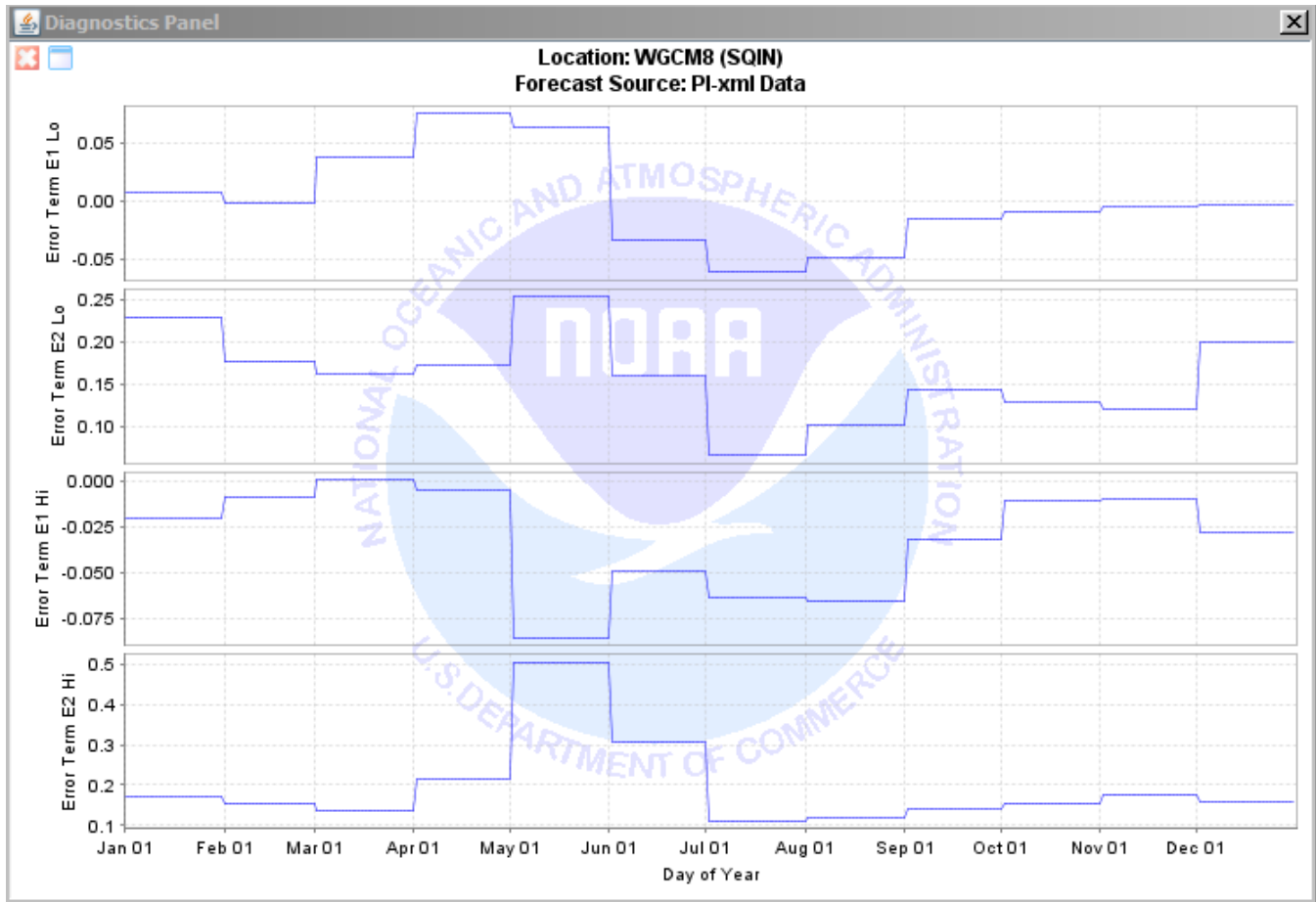
Again, the impact of snowmelt is clear on the error terms, but it is generally harder to interpret the error values beyond any seasonal pattern.

The cumulative distribution functions (CDFs) employed in the normal quantile transform to convert streamflow to and from normal space should also be examined. These CDFs should be similar for months that will be grouped into seasons. Here are the CDFs for the month of January:
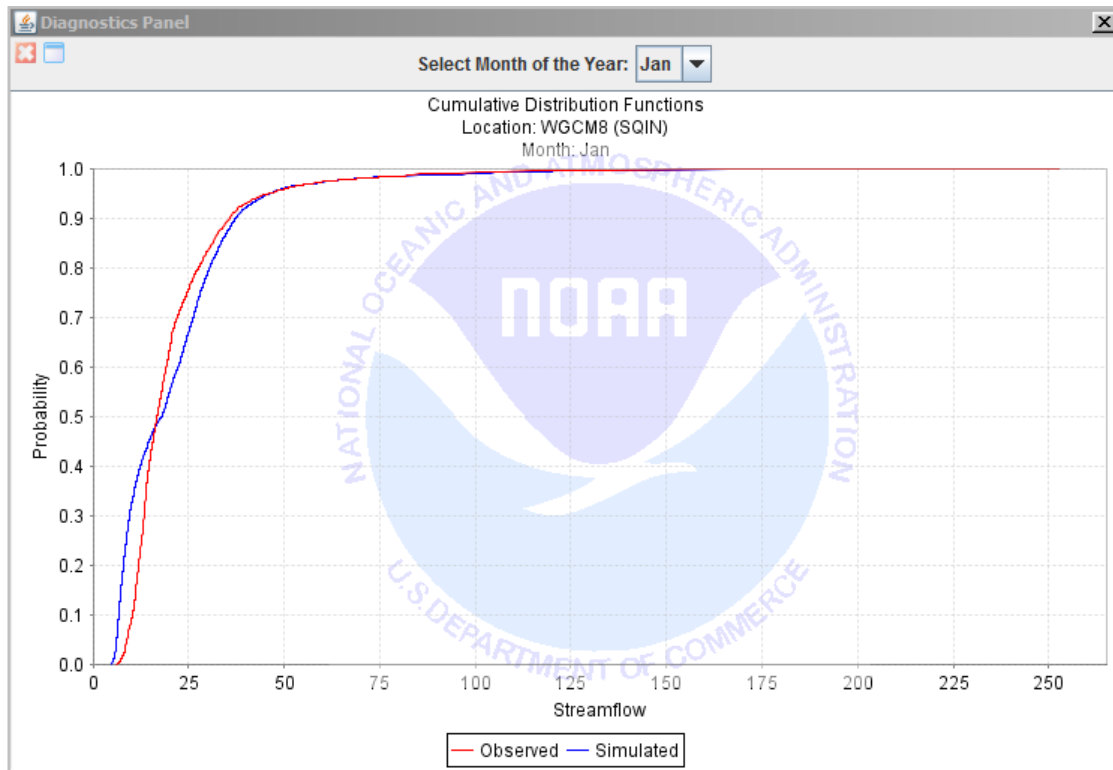
Here are the CDFs for the month of June:
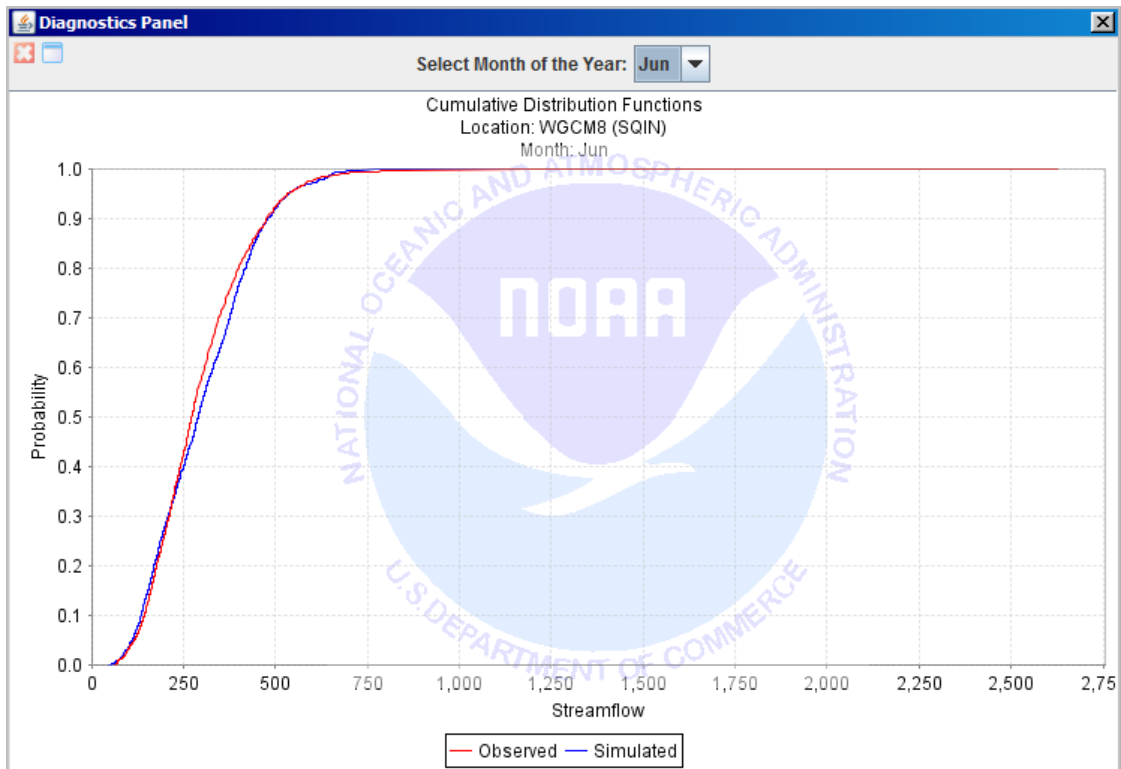
The other months are not shown for this example.

For this segment, then, a single season spanning months not impacted by snow melt is likely sufficient, with one or two seasons employed during the snowmelt months. The lengths of the snowmelt seasons should be determined not only from these diagnostics, but through leveraging any local-knowledge of the basin conditions and considering the length of the historical data available for parameter estimation.

It is unlikely to be practicable to perform the above analysis for individual segments. Instead, wherever possible, a region or RFC should be grouped into classes of basin that are likely to exhibit similar streamflow climatologies and model error characteristics and the above analysis conducted for a representative subset of basins from each group.

2.1.4 High Flow/Low Flow Threshold

The EnsPostPE allows for the use of a single threshold, defined by a climatological probability, to demarcate the sample of historical simulations and corresponding observations into two categories based on the simulated flow: high flows (above the threshold) and low flows (below the threshold). Parameters of the error model are estimated independently for each category, though they make use of a single pair of CDFs, one for historical simulations and one for observations, spanning both categories of flow. The default threshold probability is 0.5 (the median or 50th percentile) so that approximately half of all samples will be employed when estimating parameters for each category of flow. Without an explicit evaluation of this threshold probability (e.g. through hindcasting and validation), the default threshold should be used. When the threshold probability is modified, it will adjust the sample size available for estimating the parameters in each category (above and below the threshold), with one using more samples and the other using fewer samples. This may require a longer record of historical simulations and observations, in order to produce reliable estimates of the parameter values for the category with fewer samples.

2.2 Post-processing Ensemble Forecasts with the EnsPost

Ensemble forecasts are post-processed with the HEFSEnsPostModelAdapter, which is configured as a GeneralAdapter in CHPS; see Section 5.1. Input data required for execution of the adapter consists of a single time series of observed streamflow from which the latest observed value is acquired (or computed if aggregation is necessary), and the streamflow ensemble forecast to post-process. The input data is exported from CHPS and the resulting post-processed ensemble generated by the HEFSEnsPostModelAdapter is imported into CHPS. The EnsPost can be executed for a single segment or multiple segments at once (i.e., configured in a single module).

The ensemble forecasts input to the HEFSEnsPostModelAdapter should include the same sources of error and bias that are represented in the historical simulations from which the EnsPost parameters are estimated. This is particularly important when the EnsPost is applied at downstream locations. Examples of potential inconsistencies between the historical and operational data are considered below, together with recommendations on how to avoid or minimize these inconsistencies.

2.2.1 Use of raw (unadjusted) operational forecasts

Summary of Recommendation: The EnsPost should be applied to the raw (unadjusted) streamflow forecasts at each downstream location.

The EnsPost is calibrated for each location, separately, using raw hydrologic simulations. Thus, in order to account for the combined uncertainties in the hydrologic modeling from upstream (routed) and downstream (local) flows, the EnsPost should be applied, operationally, to the raw streamflow forecasts at each downstream location (i.e. the local contribution plus the raw contribution from upstream).

2.2.2 Runtime Modifications

Summary of Recommendation: Avoid applying runtime modifications that systematically adjust the raw streamflow forecasts when the EnsPost is likely to (partially or completely) duplicate these adjustments.

Systematic adjustments to the operational streamflow forecasts via runtime modifications may be partially or fully repeated by the EnsPost. These runtime modifications include modifications to hydrologic model states (e.g., snow states or soil moisture states) or modifications that introduce sources or sinks of flow, such as streamflow diversions. Such "double accounting" should be avoided, wherever possible. However, runtime modifications that involve non-systematic ("event-based") adjustments that are not (or only weakly) reflected in the EnsPost parameters might reasonably be applied, operationally.

Clearly, if runtime modifications are not reflected in the historical simulations from which the parameters of the EnsPost are estimated and the EnsPost is subsequently applied to operational streamflow forecasts that *do* include those runtime modifications, an inconsistency emerges between the EnsPost parameters and the biases that are present in the operational forecasts. Such inconsistencies are inevitable when the runtime modifications involve adjustments to the hydrologic model states, because the EnsPost is applied to hydrologic forecasts that are derived from those states. In contrast, runtime modifications that are applied to the (post-processed) streamflow forecasts themselves should not introduce any inconsistencies, because the EnsPost has been applied before any runtime modifications, in keeping with the EnsPost parameters. However, adjustments of the operational forecasts are (much) less common than adjustments of the hydrologic model states.

It is difficult to generalize about the relative impacts of avoiding inconsistencies between the EnsPost parameters and operational forecasts, by omitting runtime modifications, and the consequences of ignoring important operational datasets (e.g. on snow states). Nevertheless, some generalizations are possible. First, emphasis should be placed on minimizing these inconsistencies wherever possible (e.g., by replacing runtime modifications with rules that can be applied historically). Second, runtime modifications that involve "event-based" (non-systematic) adjustments are less likely to introduce material inconsistencies between the EnsPost parameters and the operational forecasts. Finally, where inconsistencies do emerge, the operational forecasts should be archived, monitored, and validated once the archive is sufficiently large (although this implies several years of data), in order to identify the impact of those inconsistencies.

2.2.3 Order of Forecast Operations

Summary of Recommendation: Forecast operations that cannot be reproduced in the historical simulations should be applied <u>after</u> the EnsPost.

Streamflow operations that are applied to the operational forecasts, but not to the historical simulations, should be applied *after* the EnsPost. Examples include blending and ADJUST-Q. Since the parameters of the EnsPost will account for biases that are present before blending/ADJUST-Q, any corrections made by the EnsPost to operational forecasts that include blending/ADJUST-Q may be duplicative or otherwise erroneous. Also, since the EnsPost is an autoregressive model, it already incorporates the lagged observation when adjusting the operational forecast. Thus, even when the EnsPost is applied correctly (before any blending), the subsequent blending operation may be duplicative (this should be tested through hindcasting and validation).

2.3 Regulated Rivers

Summary of Recommendation: Initial applications of the EnsPost should focus on unregulated rivers, natural flows at regulated locations, or locations for which upstream regulations are minimal.

Ultimately, the EnsPost will be applied at regulated and unregulated locations alike. Indeed, the hydrologic errors and biases typically account for a large fraction of the overall uncertainties in streamflow forecasting, and the EnsPost may (in some cases) allow for the modelling of regulation uncertainties. However, there are several challenges for the successful application of the EnsPost at regulated locations. Until then, some general recommendations are possible:

- Consistency in the calibration and operational use of the EnsPost is critical, both in regulated and unregulated rivers. In other words, the historical simulation errors must be consistent with those observed operationally (see above). For example, it would not be appropriate to calibrate the EnsPost on unregulated flows (e.g., ignoring a diversion) and then apply the EnsPost, operationally, to regulated flows (e.g., with a diversion introduced as a runtime modification). Inconsistent applications of the EnsPost may increase (rather than reduce) the errors and biases in the operational streamflow forecasts;
- In the absence of an explicit regulation model (e.g., HEC-ResSim) or a long historical archive of operational adjustments, there is limited scope to apply to the EnsPost to regulated locations, as the EnsPost must be calibrated (and validated) with historical data;
- The modeling errors associated with river regulations, and particularly those involving reservoir operations, may undermine the statistical assumptions adopted by the EnsPost. For example, the EnsPost does not explicitly model timing or phase errors and assumes that any seasonal variations are reasonably aligned with calendar months. The consequences of undermining these assumptions are difficult to predict. For example, in some cases, the EnsPost may operate imperfectly and still increase forecast skill. Some examples are provided in Section 3.3; and
- For the above reasons (among others), hindcasting and validation are central to the application of the EnsPost in regulated rivers. Such applications are likely to involve several judgements and trade-offs, such as the length and consistency of the historical record, the choice of seasonal and threshold

stratification, the consistency of the historical and operational data, and whether the modeling should consider naturalized or regulated flows. These judgements must be tested through validation and, where necessary, revised.

The OWP is currently conducting hindcasting and validation of the EnsPost in regulated rivers, in order to provide guidance on how best to calibrate and apply the EnsPost for different types of regulations. Upon completion, this manual will be updated with any new guidance.

2.4   Frequently Asked Questions

In the following sections, frequently asked questions will be addressed. Instructions for configuration will not be provided here, explicitly, but, rather, described at a high level with more detailed descriptions in this manual or the configuration guides referenced as required.

2.4.1 How do I add new segments?

After the parameters for the initial set of EnsPost segments are estimated and the EnsPost configured, it will most likely be necessary to add additional EnsPost segments. This topic is covered in each of the configuration guides:

- *EnsPostPE Configuration Guide*, Section 4
- *EnsPost Configuration Guide*, Section 3

When adding new segments, follow the instructions provided in each of the configuration guides in the order listed above. After configuring EnsPostPE, estimate the parameters before moving on to configuring EnsPost.

3   Science of EnsPost

3.1   Introduction

The HEFS produces ensemble streamflow forecasts out to a year into the future. Unlike a single-valued, or deterministic, forecast, an ensemble forecast provides an estimate of the forecast uncertainty, i.e., predictive uncertainty. Such an ensemble forecast, from which various probabilistic forecasts may be derived, should account for various sources of uncertainty in the forecast process. Toward that end, it is useful to examine briefly how hydrologic forecasting is done at the River Forecast Centers (RFC).

A collection of hydrologic models (i.e., SAC, SNOW-17, etc.) is supplied with a forecast of the input forcings, the initial conditions, and recently observed flow, and run over a specific forecast horizon. The hydrologic models solve many (complex) equations that approximate the various, generally nonlinear processes between the forcing variables, i.e., the input, and streamflow, i.e., the output. This output contains uncertainties propagated from those in the input forcing forecast, initial conditions, hydrologic model parameters and structures, and observations of flow, and other variables. Because of the complex, multiscale, nonlinear dynamics involved, it is not an easy task, if at all feasible, to isolate uncertainty contributions from among the different sources.

Broadly, these uncertainties may be grouped into two categories, input uncertainty and hydrologic uncertainty. The former comprises uncertainties in the forecast of the input variables to hydrologic models, such as precipitation and temperature. The latter comprises uncertainties in the initial conditions, parameters and structures of the hydrologic models, and those due to human influences such as flow regulations. As part of the HEFS, the forecast input uncertainties are modeled by the Meteorological Ensemble Forecast Processor (MEFP), whereas hydrologic uncertainty is modeled by the hydrologic ensemble post-processor, EnsPost. Figure 2 illustrates how the input and hydrologic uncertainties, as represented by ensembles, are propagated and integrated to represent the total uncertainty.

3.2   Methodology

A number of different ensemble post-processing techniques have been developed to model the hydrologic uncertainty. There are, however, only a few that have demonstrated applicability in an operational environment. The technique developed at the Office of Water Prediction (OWP) is EnsPost. EnsPost was developed with the intention of being simple, parsimonious (in the sense that it involves a minimal number of parameters), relatively easy to understand, and not very computing resource-intensive. Similarly to other statistical techniques, however, EnsPost requires a long period of record to estimate the parameters reliably. Because the technique is designed to model the hydrologic uncertainty only, EnsPost uses simulated (i.e. forecast streamflow with perfect future input forcing) rather than forecast streamflow (see Figure 2).

In the context of streamflow simulation, quantifying hydrologic uncertainty amounts to quantifying the error in the model simulated flow relative to the verifying observed flow. The magnitude of this error usually depends on that of the simulated flow. When this error is added to the simulated flow, the sum represents the error-corrected flow. Hence, if this error is modeled as an uncertain, or random, variable, the sum of this random error and the simulated flow represents the simulated flow that reflects
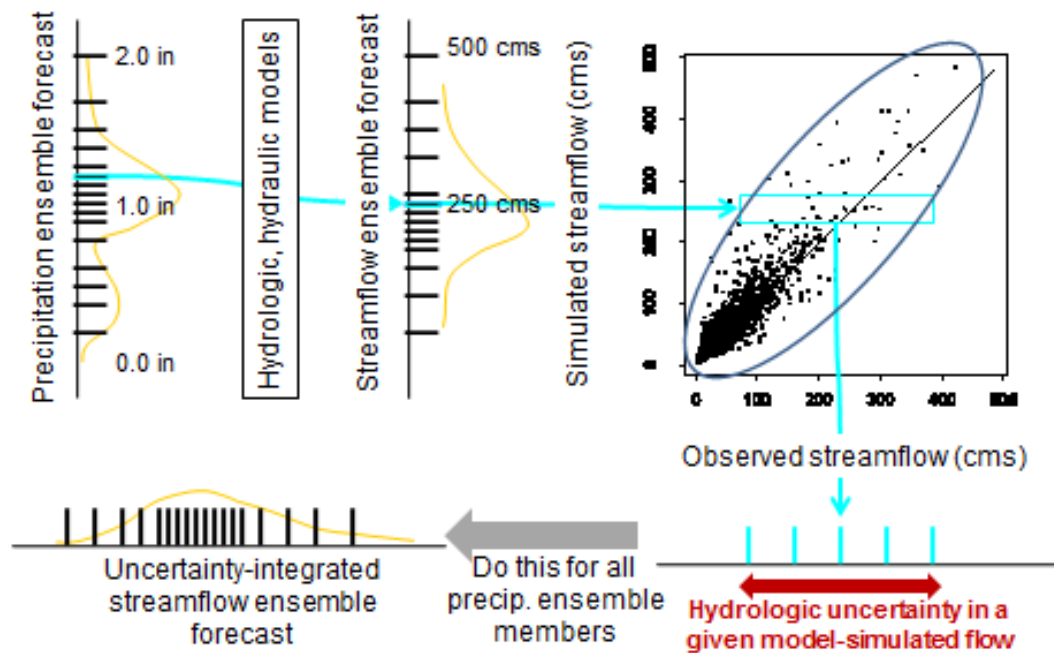
Figure 2: Illustration of integration of input and hydrologic uncertainties in hydrologic ensemble forecasting

hydrologic uncertainty. In general, the statistical properties of this error depend on the magnitude of the simulated flow. As such, one may consider EnsPost as estimating the conditional probability distribution of observed flow (i.e. the error plus the simulated flow) given the simulated streamflow (see Figure 3b).

The conditional distribution of observed flow may be estimated from the joint distribution of the observed and simulated streamflows. There are a number of different ways to estimate the joint distribution (e.g., Krzysztofowicz, 1999, Montanari and Brath, 2004, Seo et al., 2006, Chen and Yu, 2007, Hantush and Kalin, 2008, Montanari and Gross, 2008, Todini, 2008, Bogner and Pappenberger, 2011, Brown and Seo, 2012). It is well-known that variables such as precipitation and streamflow are, in general, skewed. This makes modeling of the joint distribution rather difficult. For that reason, in EnsPost, these variables are normal-transformed so that the transformed variables are individually standard normal. Such transformation is referred to as Normal Quantile Transform (NQT). Figure 3a illustrates NQT for observed and simulated flows in which the scatter plot in the upper-left corner represents the scatter plot in the upper-right corner of Figure 3 (but with x- and y-axis reversed). The NQT is popular and has been widely used in hydrologic and related applications over the years.

EnsPost transforms both the observed and simulated flows via NQT (Figure 3a), and then estimates the conditional probability distribution of the (to-be-realized) observed flow given the simulated flow (Figure 3b) and the most recently observed flow via linear regression in the normal space. The particular regression (or time series) model used in EnsPost is called the first-order autoregressive model with an exogenous input, or ARX(1,1) (Box and Jenkins, 1976). The predictors of the model are the model-simulated flow and the most recently observed streamflow. The predictand of the model is the (to-be-realized) observed flow valid at the same time as the model-simulated flow. Additional details may be found in Seo et al. (2006). The equation for the model is as follows:
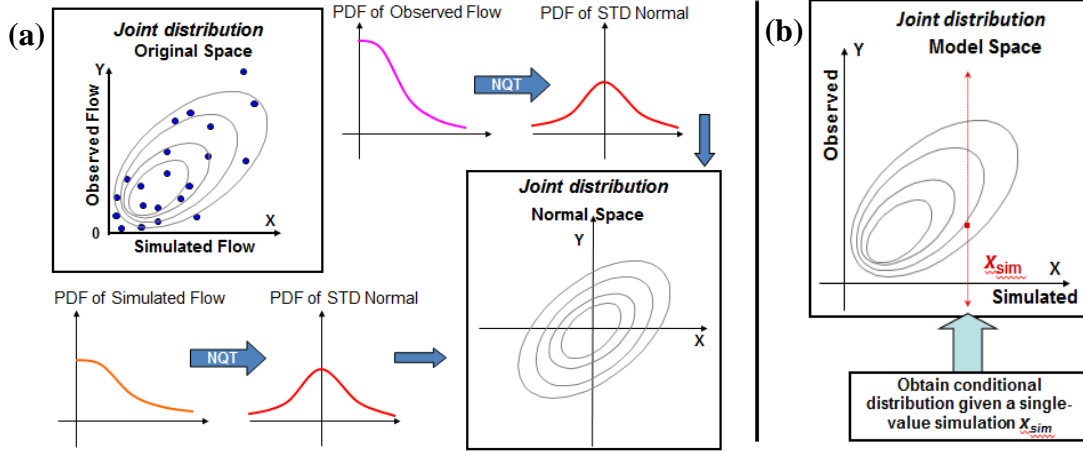
Figure 3: Illustrations of a) normal quantile transform (NQT) and b) conditional probability distribution.

$$Z_{o,k+1} = (1 - b)Z_{o,k} + bZ_{s,k+1} + E_{k+1} \tag{1}$$

where $Z_{o,k}$ and $Z_{o,k+1}$ denote the normalized observed flows at time steps k and k+1, respectively, $Z_{s,k+1}$ denotes the normalized model-predicted flow at time step k+1, E_{k+1} denotes the random error representing the hydrologic uncertainty at time step k+1 in the normal space, and b denotes the weight given to the normalized model prediction ($0 \leq b \leq 1$).

The regression model, ARX(1,1), and residual error model, AR(1), are calibrated using the historical time series of observed flow and the corresponding model (historical) simulated flow. The statistical properties of observed flow and the error in the model simulation vary according to season and the magnitude of flow. As such, the NQT curves and the regression parameters are stratified according to season and magnitude of simulated flow. In EnsPost, the user may define different levels of seasonal stratification (e.g., biannual, 4-seasonal or monthly) and choose a threshold flow (a probability quantile that defines the threshold) in order to stratify the regression parameters according to the magnitude of the simulated flow. The level of seasonal stratification and the threshold flow are selected such that, in each category, nonstationarity and heteroscedasticity are reduced as much as possible so that the magnitude of variability in streamflow does not vary too much in time (i.e. reasonably stationary) or depend too much on the magnitude of streamflow (i.e. reasonably homoscedastic), and different categories capture disparate temporal correlation structures (e.g., very fast/slowly-decaying serial correlation in high/low flows). In practice, however, the period of available recorded data may not be large enough to allow monthly, or even 4-seasonal, stratification.

The parameters of the ARX(1) model are optimized by minimizing the mean continuous ranked probability score (CRPS, Hersbach, 2000) of the post-processed streamflow ensembles. The parameters of the AR(1) model for the residual errors are then computed accordingly. The CRPS is one of the most widely used performance measures in ensemble verification and reflects multiple attributes, including reliability and resolution. (Estimation options also allow for root mean-squared error (RMSE) of estimated flow and root mean-squared error of quantile-quantile plot points (QRMSE) to be incorporated in the error measure to minimize.)
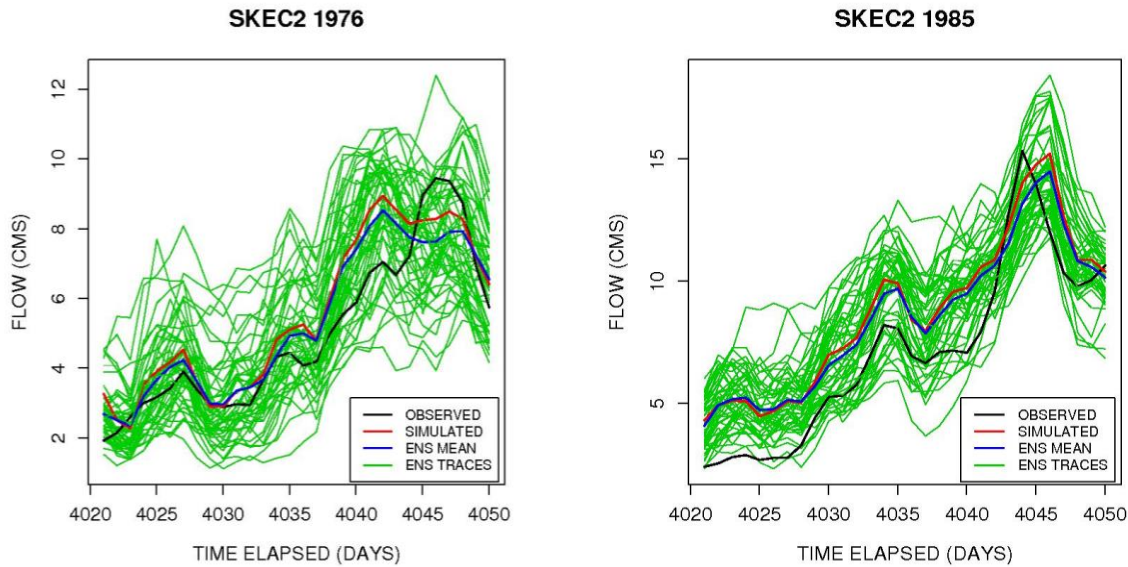
Figure 4: Example streamflow ensembles from EnsPost. The plot consists of observed flows (black), simulated streamflows, ensemble traces (green), and ensemble mean (blue); the ensemble mean is simply the arithmetic average of the ensemble traces of streamflow

Ideally, one would like to see the error in the model simulation to be completely random (i.e. white-noise). In reality, however, the above error is very often correlated in time. Such non-white-noise error structures arise because ARX(1,1) is a very simple model and can only capture the first-order autoregressive (i.e. Markovian) behaviors of the error. In EnsPost, this temporal dependence of the error is modeled as first-order autoregressive (AR(1)), the details of which may be found in Regonda et al. (2012). Figure 4 shows examples of streamflow ensembles from EnsPost for the Snake River near Montezuma (SKEC2) in the CBRFC's service area. Note that ARX(1,1) and AR(1) do a reasonably good job of reproducing the noisiness and temporal pattern of variability present in the observed flow.

3.3 Assumption and Limitations

EnsPost is a purely statistical technique that combines probability matching and linear regression (albeit in normal space). It assumes that the statistical relationships estimated from the historical data stay the same. If the climatological distribution of the observed flow changes due, e.g., to climate change or urbanization, the above assumption no longer holds. The regression model used in EnsPost is parsimonious (it has only a few parameters) and hence does not require a large amount of data. Probability matching, on the other hand, requires reliable estimation of the empirical cumulative probability distribution functions (CDF) particularly in the all-important upper tail of the distribution and is very data-intensive. Experience so far has shown that at least 20 years' worth of data is necessary to obtain reasonably reliable parameters with 2-season (wet and dry) stratification.

When the above assumptions and general requirements are met, EnsPost performs as designed. Figure 5 shows examples of the model-simulated flow vs. the verifying observed flow for Lake Mendocino (LAMC0, 1962-2002) and Hopeland (HOPC1, 1961-2004) in the CNRFC's service area. Also shown in the right-hand side of the figure are those of the cross-validated ensemble mean flow from EnsPost vs. the verifying observed flow. The ensemble mean is simply the arithmetic average of the ensemble traces of streamflow obtained from EnsPost by post-processing the model-simulated flow. Note that EnsPost
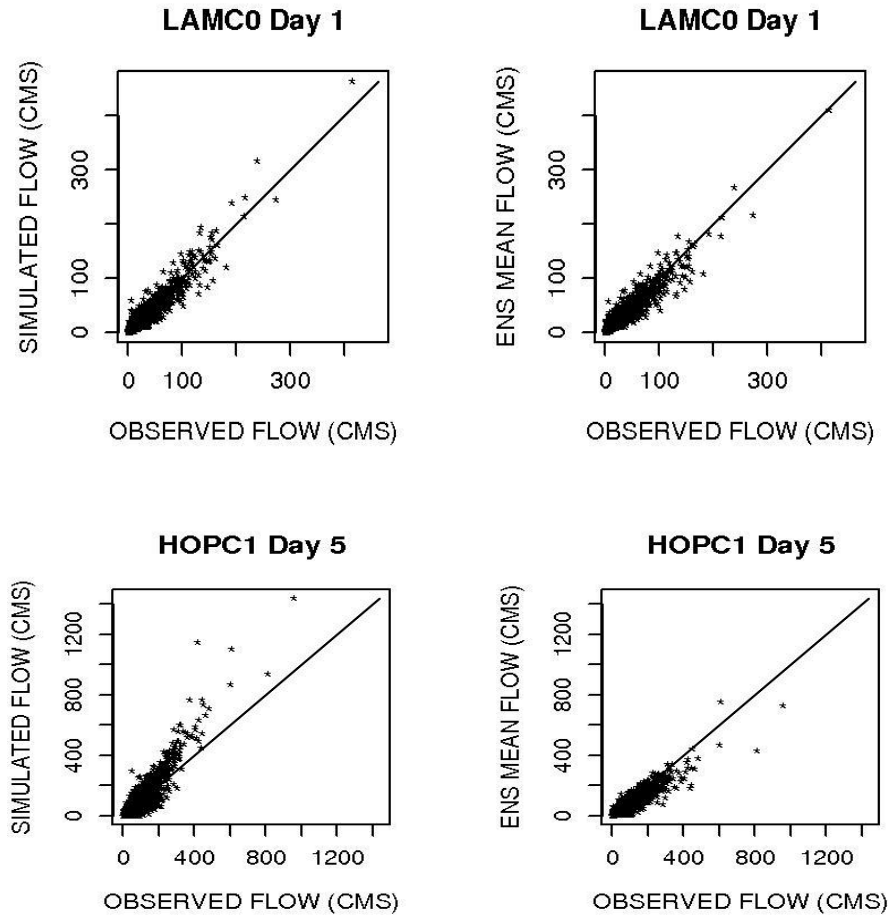
Figure 5: (left) Examples of model-simulated flow vs. verifying observation and (right) ensemble mean flow from EnsPost vs. verifying observed observation.

successfully corrects the very small bias in the LAMC0 simulation (but little or no improvement otherwise) and the rather large bias in the HOPC1 simulation, resulting in substantial improvement over the model simulation before post processing.

The statistical properties of the simulation error depend greatly on the streamflow generation mechanism. For example, simulation of rain-on-snow events tends to have larger errors. Modeling such storm type-dependent errors reliably, however, requires a much larger amount of training data, which is not likely to be available. As such, great care should be taken in applying EnsPost to model simulations with disparate errors. Figure 6a shows the model-simulated flow vs. the verifying observation for Saxton (SAXP1) in the Juniata River Basin in the MARFC's service area. Figure 6b shows the corresponding cross-validated ensemble mean flow from EnsPost vs. the verifying observation. The parameter estimation period coincided with the API calibration period of Sep 1963-Jan 1974. The annotated data points are associated with rain-on-snow events in 1979 and 1996. Note that EnsPost is not successful in reducing the errors in these outlying model simulations.

EnsPost does not explicitly consider timing, or phase, errors. As such, in the presence of significant timing errors, the post-processes ensemble traces may not be very realistic. Finally, EnsPost assumes that streamflow has a degree of predictability as expressed by serial correlation, and that the model
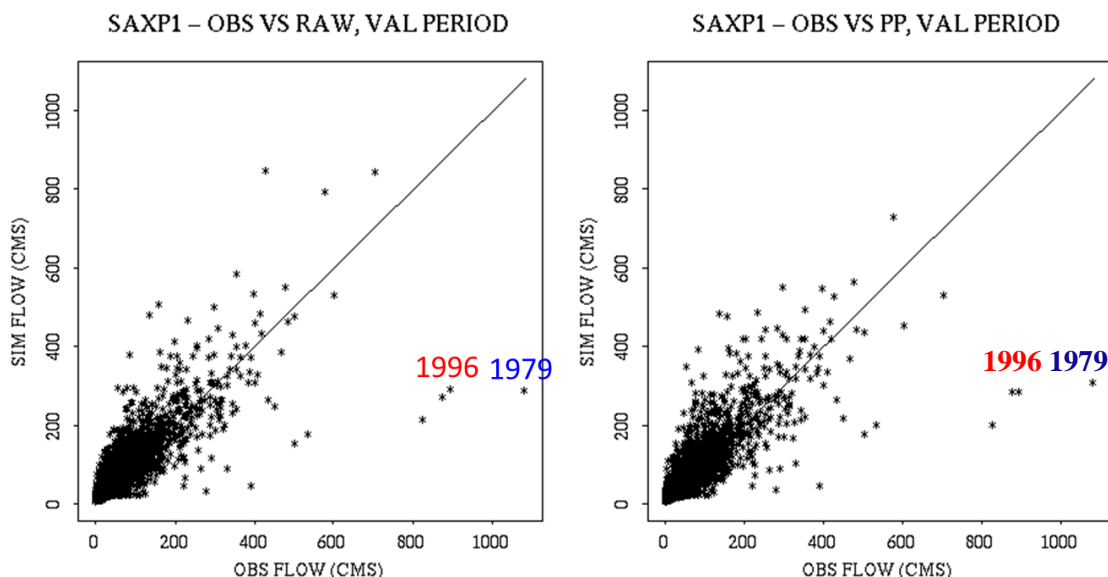
Figure 6: a) Model-simulated flow vs. verifying observed flow for Saxton, PA, and b) the corresponding ensemble mean flow from EnsPost vs. the verifying observed flow. The annotated data points are rain-on-snow events.

simulation is skillful. For regulated flow, however, the above assumptions may not hold and hence EnsPost may be of very limited utility. Figure 7a shows the model-simulated flow vs. the verifying observation for Raystown Dam on the Raystown Branch of the Juniata River in the MARFC's forecast area. Figure 7b shows the corresponding ensemble mean flow from EnsPost vs. the verifying observation. The ensemble mean results are from parameter estimation and hence offer an assessment of the goodness of the statistical model used in EnsPost for dealing with regulated flows. Note that, while EnsPost reduces the very large errors associated with regulated flows, it does so only at the expense of introducing a large bias to the overall results. Figure 8shows examples of the ensemble traces generated by EnsPost when the observed flow is subject to regulation. Note in the figure that EnsPost is largely unable to capture the unnatural temporal patterns in the observed flow associated with regulations.

Currently, the EnsPost operates at either a 6-hourly or daily time step (user selectable). Once the calibration time step has been selected, the EnsPost will aggregate the input time series as needed (e.g., either hourly to 6-hourly or either hourly or six-hourly to daily), both for calibration and operational application. Following any required aggregation, the time series of observations and simulations/forecasts must be equivalent quantities (e.g. instantaneous), and not simply available at equivalent time steps. For example, when estimating parameters at a 6h timestep, it must be possible to derive 6h instantaneous flows (QIN/SQIN) or 6h average flows (QME/SQME) from the available inputs, either by pairing them directly or aggregating one or both of them, and the same quantities must be available operationally. In this case, 1h QIN and 6h SQME would lead to a 6h calibration whereby the 1h QIN is averaged to produce 6h QME. However, if 6h QME and 6h SQIN were provided as input, then the EnsPostPE would produce an error, indicating that the two quantities are incompatible. When the observations and simulations/forecasts are all available at a 1h time step, and a 6h calibration is requested, the inputs are always aggregated to 6h averages, rather than using the instantaneous 6h quantities.

Operationally, the post-processed streamflow time series may, as a run-time option (see disaggOutput property described in Section 5.1.3), be disaggregated or, more precisely, interpolated to the time step of the original, raw ensemble. For example, if the calibration time step for a set of parameters is daily, the EnsPost executes at a daily time step. If the ensemble to post-process is 6-hourly, the EnsPost can interpolate the daily results to a 6-hourly time step, if the user specifies the appropriate run-time option. The interpolation algorithm is based on a constrained polynomial spline (source: http://www.korf.co.uk/spline.pdf), which provides a compromise between smoothness and physical realism (e.g. it will not produce negative streamflows). The interpolation uses the observed streamflow at the forecast issuance time as the left boundary. A further post-processing step ensures that the average of the interpolated values is equal to the aggregated input for the corresponding time period; in other words, that the bias-correction is preserved at the timestep used to calibrate the EnsPost. Thus, when the EnsPost is calibrated at a 24h scale (24h QME/SQME) and the outputs are interpolated to a 6h scale (6h QME/SQME), the average of the 6h values over the corresponding 24h period will be equivalent to the 24h output. However, there is no downscaling involved, only interpolation.

If the EnsPost outputs are desired at a particular timestep for operational purposes (e.g. 6h), the archived data may support a choice between calibrating the EnsPost at an aggregated timestep (e.g. 24h) and interpolating the results to a finer timestep (e.g. 6h) or calibrating the EnsPost at the finer timestep (6h), without any interpolation. The expected outputs from the EnsPost are summarized below for the different input time scales and calibration options prior to any interpolation.

Table 1: Expected outputs from the EnsPost for different input time series and calibration time steps. **Note that providing any input series at a time step coarser than the calibration time step will result in an ERROR.**

Observations	Simulations/forecasts	Calibration time step	EnsPost
1h QIN	1h SQIN	6h	6h SQME
1h QIN	6h SQIN	6h	6h SQIN
1h QIN	6h SQME	6h	6h SQME
6h QIN	1h SQIN	6h	6h SQIN
6h QIN	6h SQIN	6h	6h SQIN
6h QIN	6h SQME	6h	ERROR
6h QME	1h SQIN	6h	6h SQME
6h QME	6h SQIN	6h	ERROR
6h QME	6h SQME	6h	6h SQME
1h QIN	1h SQIN	24h	24h SQME
1h QIN	6h SQIN	24h	24h SQME
1h QIN	6h SQME	24h	24h SQME
1h QIN	24h SQME	24h	24h SQME
6h QIN	1h SQIN	24h	24h SQME
6h QIN	6h SQIN	24h	24h SQME
6h QIN	6h SQME	24h	24h SQME
6h QIN	24h SQME	24h	24h SQME
6h QME	1h SQIN	24h	24h SQME
6h QME	6h SQIN	24h	24h SQME
6h QME	6h SQME	24h	24h SQME
6h QME	24h SQME	24h	24h SQME
24h QME	1h SQIN	24h	24h SQME

| 24h QME	6h SQIN	24h	24h SQME
24h QME	6h SQME	24h	24h SQME
24h QME	24h SQME	24h	24h SQME

The choice between calibrating the EnsPost at an aggregated time scale, followed by interpolation, versus calibration at a finer time scale, should be driven by the anticipated forecast quality at the timestep of greatest interest for operational use and, therefore, guided by hindcasting and validation. In the absence of hindcasting and validation, it will be difficult to determine whether a 6h calibration or a 24h calibration should be preferred if the outputs are required at a 6h (or finer) scale. Indeed, several factors will influence the quality of the EnsPost outputs at a 6h or finer scale, such as the period of record for which the inputs are available (potentially shorter for 6h or 1h data than 24h data) and the basin conditions (a 24h calibration followed by interpolation is less likely to perform well in flashy basins). Based on limited hindcasting and validation by the developers of EnsPost, an improved calibration is typically achieved when using a longer period of aggregated inputs (24h QME/SQME), rather than a shorter period of instantaneous inputs. However, in flashier basins and when the inputs are available for an equivalent period (e.g. 10+ years), a 6h calibration may lead to improved outputs at a 6h (or finer) time step than a 24h calibration followed by interpolation. When the EnsPost is calibrated with aggregated flows, rather than instantaneous quantities, the interpolated results are unlikely to reproduce the temporal patterns in the observations at the desired output scale (particularly in flashy basins), but they will be smooth and "visually realistic", and they may produce better verification results (e.g. because the aggregated time series comprise a longer period and are easier to model with the EnsPost). Thus, hindcasting and validation is strongly recommended but, in the absence of this, calibration at a 24h scale (24h QME/SQME), followed by interpolation to the desired model timestep is more likely to produce better results.
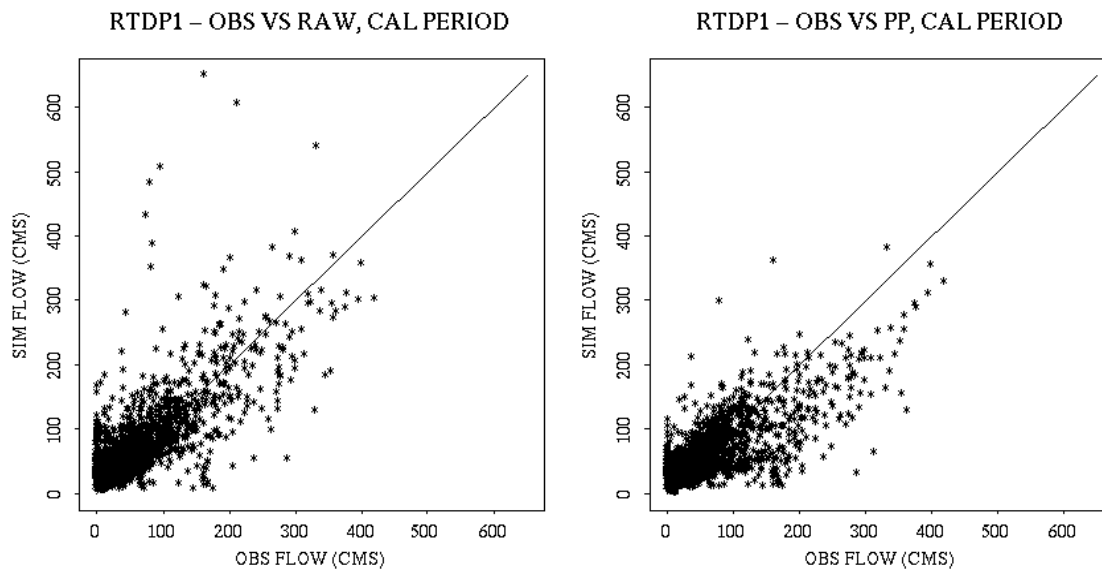
Figure 7: a) Model-simulated flow vs. verifying observed flow for Raystown Dam, PA, and b) the corresponding ensemble mean flow from EnsPost vs. the verifying observed flow.
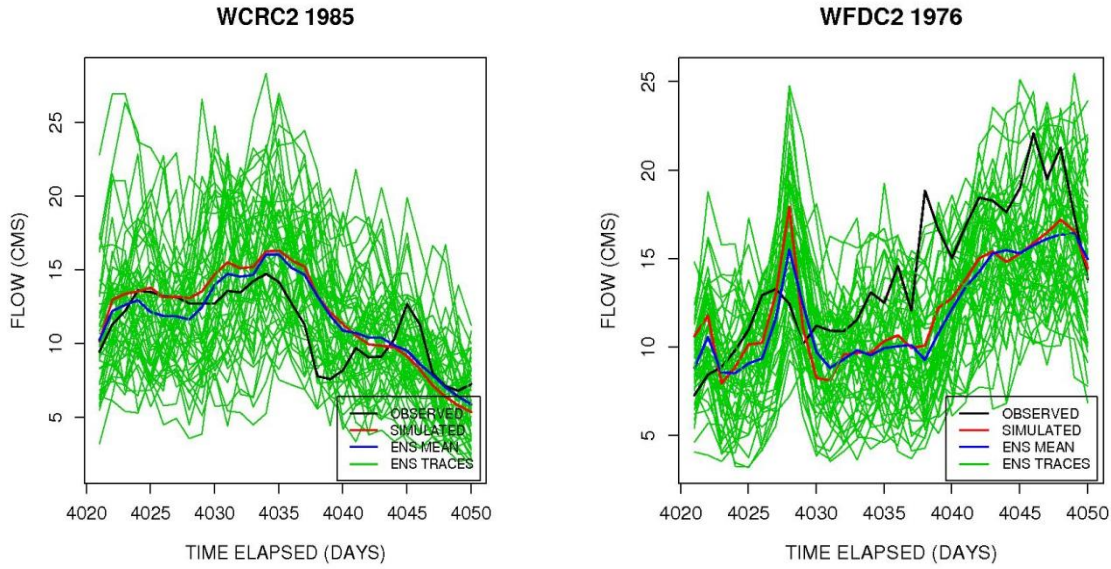
Figure 8: Example streamflow ensembles from EnsPost for regulated flows: a) Willow Creek Reservoir Near Granby, CO, and b) Williams Fork Reservoir near Parshall, CO.

3.3.1 Applying for a downstream location

For a downstream location, the model simulation that contains hydrologic uncertainty relative to the verifying streamflow observation at that location is the combined flow. As such, the sum of the routed flow from upstream and the local flow from downstream should be post-processed, rather than the local flow. However, the routed flow from upstream must not include the impact of EnsPost (i.e., it must be "raw") since the historical simulation will not include the impact of EnsPost.

3.4 Error models and Estimation Options

EnsPost includes three different algorithms for post-processing, referred to as *error models*. While they are described as different, they belong to the same family described in Section 3.2 and described mathematically in Eq. (1):

1. **Probability Matching (ER0):** Equivalent to Eq. (1) with b = 1.0 and the error term, E_{k+1}, having zero mean and variance, therefore assuming the simulation perfectly predicts the observed value after being converted into normal space so that only the cumulative distribution functions need to be applied. This is effectively equivalent to simple distribution mapping: computing the probability associated with a streamflow value via the model simulation cumulative distribution function (CDF), and then finding the streamflow value with the same probability in the observed CDF. After post-processing, the resulting ensemble will not properly quantify the hydrologic uncertainty.

2. **Error Model Deterministic (ERD):** Eq. (1) is applied as is, except that the error term, E_{k+1}, is assumed to have zero mean and variance. This will perform some additional level of debiasing, beyond just CDF matching, but after post-processing, the resulting ensemble will not properly quantify the hydrologic uncertainty.

26

3. **Error Model Stochastic (ERS)**: Incorporates all parts of Eq. (1) when applied operationally, including sampling from the error term, E_{k+1}, through the use of the AR(1) model of the residual error in order to account for hydrologic uncertainty. **(RECOMMENDED)**

Specified in parentheses is the name of the error model when referred to in the configuration of the CHPS module as a value for the run file property errorModel.

All three procedures above require a set of estimation options be specified for the calibration, or parameter estimation, of EnsPost (i.e., within the *EnsPostPE* interface). Described in Section 4.6.2, the options provide controls over the following:

Basic Options:
1. **Seasons**: Grouping months together into meteorological seasons. Typically either annual (no seasonality), biannual (wet and dry), 4-seasonal (spring, summer, fall and winter) or monthly (Jan through Dec).
2. **Calibration time step**: The time step at which parameter estimation is performed and the EnsPost will be applied; either 6-hourly or daily.

Advanced Options:
3. **Threshold**: The threshold that demarcates high and low flows for the purposes of modeling as per Section 3.2. The threshold is defined by a probability associated with the quantile to be used as the cutoff point.
4. **Omega parameter of empirical CDF**: Sets the omega parameter controlling the upper tail of the empirical cumulative distribution function beyond the largest observed value. An option exists for both the observed and simulated CDFs.
5. **Numerical integration in back-transformation**: A numerical integration is required in order to back-transform values from normal space into observed space. That integration is controlled via a lower bound (min), upper bound (max), and number of intervals defined in between.
6. **Target lead day of optimization:** The number of lead days for which to apply EnsPost in order to generate hindcasts and calculate the error measure (CRPS, by default) when optimizing its performance as part of parameter estimation (i.e., for the ERD and ERS models).
7. **Parameter Estimation/Pairs Time Step:** The time step between successive hindcasts generated by EnsPost in order to calculate that error measure. This is only applicable when a calibration time step of 6-hours is used.
8. **Number of traces/members**: The number of members to generate per hindcast in order to calculate that error measure.
9. **Weights Employed in Error Measure:** The weights to apply to the CRPS, RMSE, and QRMSE in order to compute the error measure minimized during optimization.

4   EnsPostPE Reference Manual

4.1   Overview

The EnsPostPE computes parameters used by the EnsPost to post-process streamflow ensembles operationally. Those parameters are contained in gzipped tar files (.tgz) located under the *<ens_post_root_dir>* defined within the *EnsPostPE Configuration Guide*. The EnsPostPE guides the user through a step-by-step estimation process that includes setup, estimating parameters, and accepting those parameters by placing them in the central area for use in post-processing ensemble streamflow forecasts. Alternatively, the user is also able to perform all steps in a hands-off mode via a run-all feature. EnsPostPE is a FEWS explorer plug-in, being seamlessly integrated within the CHPS/FEWS interface, and provides diagnostic capabilities to enable the user to more easily examine the historical data provided as input and the estimated parameters.

Section 4 of this manual describes how to use the EnsPostPE software interface to accomplish parameter estimation and provides details about all of the interface components. It is recommended that users read Section 4.2, *Getting Started*, prior to using the software, and refer to the other sections as needed while using the software. This manual is available via the EnsPostPE help functionality.

4.2   Getting Started

The EnsPostPE guides users through a step-by-step procedure outlined in Section 4.2.3, providing tools to allow for quality-controlling data and analyzing the parameters. This section provides basic background material pertinent to the understanding of the EnsPostPE in order to get started using the software. It explains:

1. Inputs to the EnsPostPE.
2. How to run the EnsPostPE.
3. The parameter estimation procedure through which the EnsPostPE guides the users and how that procedure connects to the interface components.
4. Core concepts for understanding and use of the EnsPostPE.

4.2.1   Inputs to the EnsPostPE

4.2.1.1   Historical Simulated and Observed Streamflow

The EnsPostPE requires historical simulated streamflow and corresponding observed streamflow.

The historical *simulated streamflow* time series is generated through the standard process in CHPS: execute the hydrologic model using historical observed forcings, including mean area precipitation (MAP), mean areal temperature (MAT), and possibly mean areal potential evapotranspiration (MAPE) data. Based on the historical data available, the model is executed starting with cold states from the beginning to the end of that record. For downstream points, the model simulation that should be used in the EnsPostPE is *combined flow*, i.e., the sum of routed flow from upstream and local flow from downstream. The combined flow simulation is used because the observed flow at a downstream point is combined flow from all upstream points.

The "warm-up period" of the historical simulation <u>must</u> be discarded before the historical simulation data is used.

> As a general rule, discard the first two years of historical simulation data before using it in the EnsPostPE. It is recommended that this be done through the use of the optional start/end date fields of the **Select Time Series Dialog** that is used by clicking on the **Export Time Series from CHPS DB Button** of the **Setup Panel** (see Section 4.5.2.1)

The historical *observed streamflow* time series is a time series of observations that corresponds to the simulated streamflow time series, being for the same gage. Those values may be recorded at a time step that differs from the time step of the historical simulation time series; see the next section.

EXPLANATION
The purpose of EnsPost is to account for hydrologic uncertainty in the hydrologic forecasts for the RFC forecast points. To quantify the hydrologic uncertainty, the EnsPost needs to be trained using the data that reflects the hydrologic uncertainty *only*; that is, data for which there is no uncertainty due to meteorological inputs for which the uncertainty is accounted elsewhere. For HEFS, that input uncertainty is accounted for by the MEFP and includes uncertainty due to future precipitation and temperature.

4.2.1.2 Time Step

Parameter estimation, or calibration, can be performed at either a 24-hour time step (default) or a 6-hour time step. However, the resulting time series, <u>after aggregation to the calibration time step,</u> must be of compatible data types. The selection of *calibration time step* is done via an estimation option described in Section 4.6.2. If the data is incompatible, the EnsPostPE will error out during parameter estimation.

For example, if parameters are estimated at a 6-hour time step, with the historical simulation provided as SQIN (instantaneous) data at a 6-hour time step, but observations provided as QME (accumulative/mean) data also at a 6-hour time step, then an error will occur. This is because the simulated data is instantaneous in nature while the observations are accumulative. See Section 3.3 for an explanation of data compatibility and Table 1 for examples illustrating data compatibility.

A STEP-BY-STEP PROCESS FOR MANUAL DATA PREPARATION
The following steps and checks can be performed to prepare the historical simulated and observed streamflow time series for use in EnsPostPE:

1. Acquire historical simulated streamflow (SQIN, QINE, or SQME) and corresponding observed streamflow (QIN or QME) time series and create separate PI-timeseries XML files for each. **Be sure to produce historical simulated streamflow time series from the same configuration that is used for operations using the CHPS/FEWS.**

2. Verify that the data in these files corresponds to the period of the record that is used for EnsPost calibration. **Be sure to discard the warm-up period for the historical simulated streamflow time series (see Section 4.2.1.1).**

3. Verify that the time system (time zone) is mentioned correctly in both files of simulated and historical observed flows. It is critical that the time system in the PI-timeseries XML files is correct. Typically, the QME is stored in Data Card format (in local time) and converted to PI-timeseries XML. In converting the files, the data must be shifted correctly and/or the time system correctly identified in the PI-timeseries XML. For the historical simulation time series, this is handled via CHPS configuration, but time system identified for the import modules must be identified properly.

4. Verify the following in the PI-timeseries XML files:

   - Location id – needs to be same in both observed and simulated data files.
   - Parameter id – for historical observed time series, it must be either QIN or QME, depending on the data type. For historical simulation time series, it must be either SQIN, QINE, or SQME.
   - Time step unit – check to ensure the time step is correctly identified.
   - Flow measurement units – check that the units are correct.

5. Develop annual hydrographs using the historical observed streamflows and identify number of seasons and months in each season.

6. Copy the PI-timeseries files of historical observed and simulated streamflow time series into the EnsPostPE run area:

   <enspostpe_run_area>/piXMLFiles

   (see the *EnsPostPE Configuration Guide*).

4.2.2 Running EnsPostPE

To use EnsPostPE, you must install it in a CHPS stand-alone as described in the *EnsPostPE Configuration Guide* and then start the CHPS session. After starting CHPS, the main toolbar will include an **EnsPostPE Button**:
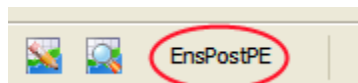
Click on this button to run the EnsPostPE. Log messages will be displayed in the standard CHPS **Logs Panel.**

4.2.3 The Parameter Estimation Procedure

The EnsPost parameter estimation step procedure is provided below. With each step, the sections describing how to use components of the EnsPostPE to perform the steps are referred to.

1. *Setup*
   Acquire historical simulated and corresponding observed streamflow time series to use from the FEWS PI-service associated with the CHPS session.

   - If the data is prepared manually (see blue inset within Section 4.2.1~~3.5.1~~), this step need not be performed.
   - Make sure to specify an appropriate data range so that the warm-up period of the historical simulation is not included; see Section 4.2.1.1.

2. *Estimate parameters*
   Specify user-defined estimation options and estimate the parameters of the EnsPost. Examine the quality of the estimated parameters to determine their acceptability.

3. *Accept (tar and gzip) parameter files*
   Create gzipped tar (.tgz) files of parameters for use in operational ensemble post-processing.

4.2.4 Core Concepts

This section discusses several concepts that are core the operations of the EnsPostPE.

4.2.4.1 The EnsPostPE Run Area

The EnsPostPE runs using files stored in the EnsPostPE run-area, as defined in the *EnsPostPE Configuration Guide* and denoted herein as *<enspostpe_run_area>*:

> *<region_dir>*/Models/hefs/hefsEnsPostPERunArea

Files stored under that directory include run-time information files, historical simulated and observed streamflow time series files, archived parameter files, and parameter files.

The user should never modify anything within the EnsPostPE run area unless specifically instructed to do so.

4.2.4.2 Run-time Information

EnsPostPE run-time information includes any information necessary for the EnsPostPE to execute and that needs to be remembered whenever the EnsPostPE is closed so that the user can pick-up where they left off upon restarting EnsPostPE. That run-time information includes the following:

- EnsPost location information
- EnsPostPE current estimation options

All other information, including the step status, is determined at run-time based on the contents of the EnsPostPE run area.

The run-time information is stored in a file underneath the system files directory within the EnsPostPE run area:

> *<enspostpe_run_area>*/.systemFiles/runTimeInformation.xml

<u>Do not modify this file unless instructed to do so by an OHD developer while debugging an issue.</u>
The file is updated once per minute while EnsPostPE is running and whenever EnsPostPE is closed.

4.2.4.3 FEWS PI-service Connection

EnsPostPE can acquire the historical simulated and observed streamflow data via the FEWS PI-service, and, in order to use the FEWS PI-service, the connection port number must be defined. After the CHPS-interface has started, check the **Logs Panel** for lines similar to the following:

> 11-04-2010 11:16:01 INFO - Started FewsPiServiceImpl on localHost : 8101
> 11-04-2010 11:16:01 WARN - Failed to start: SocketListener0@0.0.0.0:8100

The line that begins with "Started FewsPiServiceImpl…" indicates that the port number (as highlighted above) of the FEWS PI-service session initialized for the currently running session of CHPS. This is the PI-service to which the EnsPostPE should connect. If the port number is not 8100 (the default) or is not the value which was setup during installation, then EnsPostPE must be directed to the correct port number. See Section 4.5.1.1 for details on how to change the port number in EnsPostPE.

4.2.5 General Graphical User Interface Components

Some graphical user interface (GUI) components are used many times within the EnsPostPE and are described below.

4.2.5.1 Generic Summary Table

Various panels within the EnsPostPE make use of a **Generic Summary Table**, which provides information about EnsPost segments and the status of steps performed. For example:

Underneath the table is a tool bar which contains buttons that are panel specific; the example shown above is for the **Estimation Subpanel** (Section 4.4). Four buttons, however, are common to all **Generic Summary Tables:**

- **Select All Button**: Selects all rows of the table.
- **Unselect All Button**: Unselects all rows, clearing the table selection.
- **Select Rows That Need Processing Button**: Select all rows for which the status in the primary status column is <u>not</u> a check mark: or . These are the rows indicating locations for which the associated step needs to be performed or updated.
- **Refresh Button**: Refresh the table, determining the status of the rows from scratch. Clicking this button is usually not necessary, but may be required if the user manually modifies files in the EnsPostPE run area.

When this table is used within a panel, it will be referred to as a **Generic Summary Table** associated with a specific step described in Section 4.2.3 and its panel specific buttons will be described.

4.2.5.2 Table Delete/Add and Status Columns

A **Generic Summary Table** within the EnsPostPE *may* include a leading column that allows for deleting or adding rows, or status columns indicating the status of steps performed. Those columns display icons as follows:

- **Delete Icon**: Click to delete the row from the associated table. Sometimes this will cause a dialog to popup requesting confirmation of the delete.
- **Add Icon**: Click to add a row to the associated table.

A **Generic Summary Table** within the EnsPostPE may include a status column related to the step in Section 4.2.3 to which the table corresponds. The status column displays the following icons:

- **Bad Status Icon**: Indicates that as step has not been performed or an error of some kind occurred while performing some other action.
- **Warning Status Icon**: Indicates that a step has been performed but needs to be updated (performed again).
- / **Good Status Icon**: Indicates that a step has been performed or some other action was successful. The icon is usually used to indicate success, but sometimes a may be used.

For all status icons, a tool tip will display further information describing the status, such as the cause of failures or why a step needs to be updated. To see the message, leave the mouse cursor over the icon without moving it for a few seconds. For example:
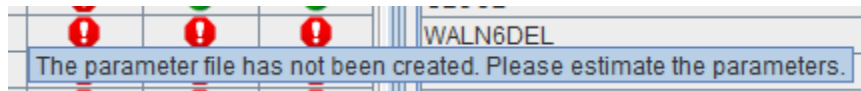
If a table within the EnsPostPE uses either a delete/add or status column, it will be stated in the description of that table. All **Generic Summary Tables** use a status column.

4.2.6 Format of the Following Sections

Sections 4.3 and 4.9 are provided as a reference for the components of the EnsPostPE interface. Each section provides the following information:

- A description of the component panel to which the section applies.
- Any special considerations required for the panel.
- A listing of the interface components, including buttons, tables, lists, etc.
- Instructions for how to perform basic tasks using the components.

4.3 EnsPostPE Main Panel

Shown in Figure 9, the EnsPostPE Main Panel is displayed as a plug-in to CHPS after initialization is completed. It includes three components:

- **Estimation Steps Panel**: Guides the user through the steps outlined in Section 4.2. A tabbed panel is provided for each of the steps.
- **Location Summary Panel**: Summarizes the status of the steps for each of the EnsPost locations. Also provides for the ability to run all steps for selected locations.
- **Diagnostics Display Panel**: Displays diagnostics that assist the user in quality controlling the data, deciding on options to use for estimation, and quality controlling and accepting the estimated parameters
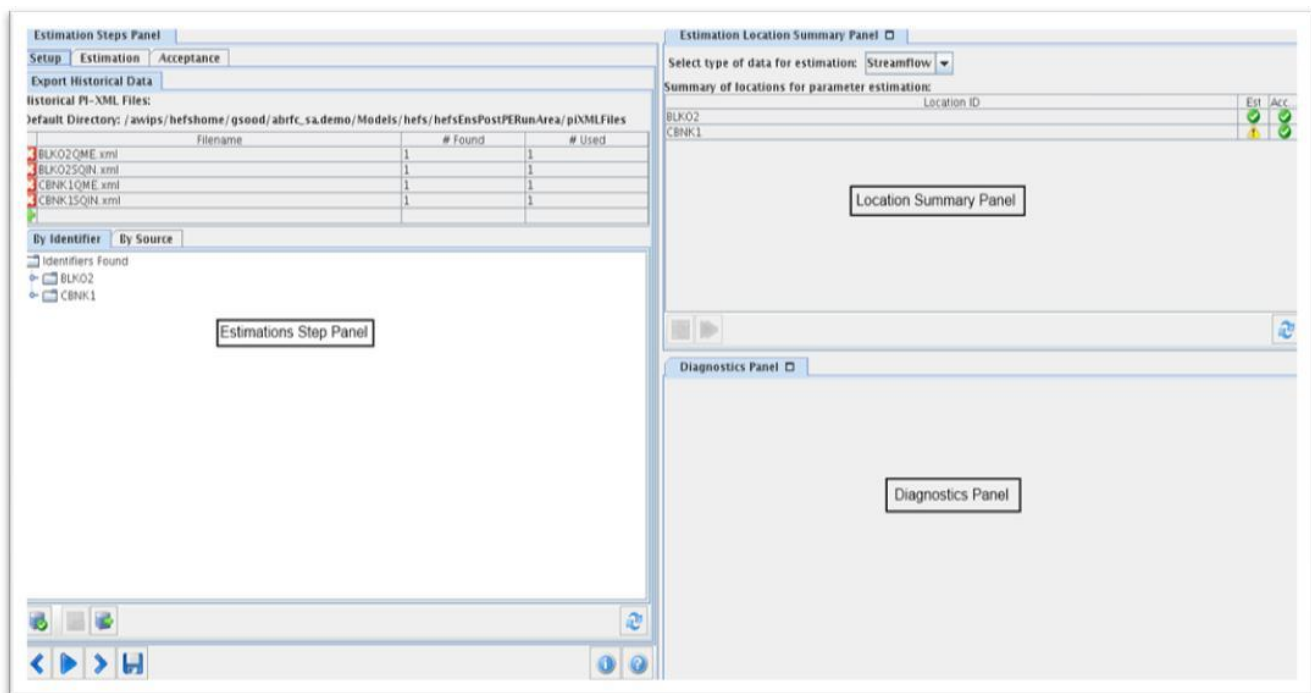
Figure 9: The EnsPostPE Main Panel, displayed upon start-up of the EnsPostPE

4.4   Estimations Steps Panel

The Estimation Steps Panel, shown in Figure 10 is positioned on the left-hand side of the **EnsPostPE Main Panel** and displays tabbed subpanels that correspond to the steps of the EnsPost parameter estimation process; see Section 4.2.3. All of the tabbed subpanels are described in sections that follow. Also provided are buttons that facilitate navigating the tabbed subpanels, an information button, and a help button.
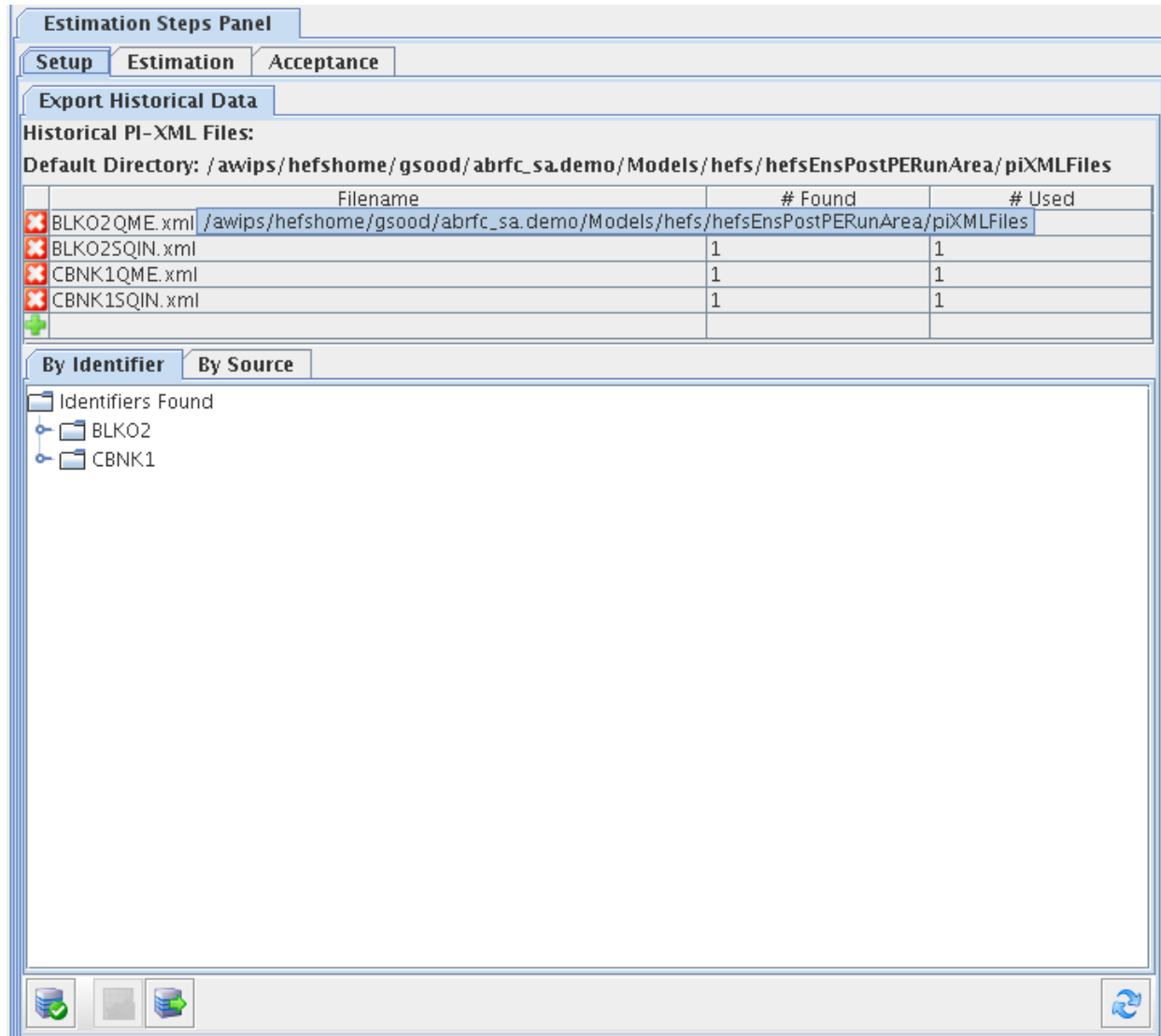
Figure 10: The Estimation Steps Panel

4.4.1 Components

The following describes the **Estimation Steps Panel** components:

- **Estimation Steps Tabbed Subpanels**: One tabbed subpanel is displayed for each of the estimation steps discussed in Section 4.2.3:

- To move between subpanels, either click on the tabs or on the **Back** and **Next** buttons described below.   A subpanel is said to be *active* if its tab is selected and its contents are currently being viewed. For example, in the image above, the **Setup Subpanel** is active.

- / **Back/Next Buttons**: Click to navigate to the previous or the next step's tabbed subpanel. The buttons are disabled if there is no previous or next subpanel.

- **Perform Step Button** (**Run Button**): Click to run the step corresponding to the active tabbed subpanel. If there is no step to perform, as for the **Setup Tabbed Subpanel**, then this button will not be present. The button is enabled only if one or more EnsPost locations for which to perform the step are selected in **Generic Summary Table** associated with the tabbed subpanel (see the description for the individual steps subpanels provided in following sections). A description of how to perform a step is presented below in Section 4.4.2.1. This button is not available in the **Setup Subpanel**.

- **Save Run-Time Information Button**: Click to force an immediate save of the run-time information. EnsPostPE saves run-time information to a file that is loaded whenever it starts, enabling it to remember user settings. The file is saved once per minute while EnsPostPE is running, when EnsPostPE is closed, and when this button is clicked.

- **About Button**: Click to display a dialog providing version information for the EnsPostPE.

- **Help Button**: Click to active help mode. When in help mode, the interface cannot be interacted with. Rather, the user can click on a component of the interface to receive help information tailored for the clicked component. The information is extracted directly from this manual and is displayed in PDF viewing dialog. The component for which help will be provided is highlighted by a faded red box; for example:
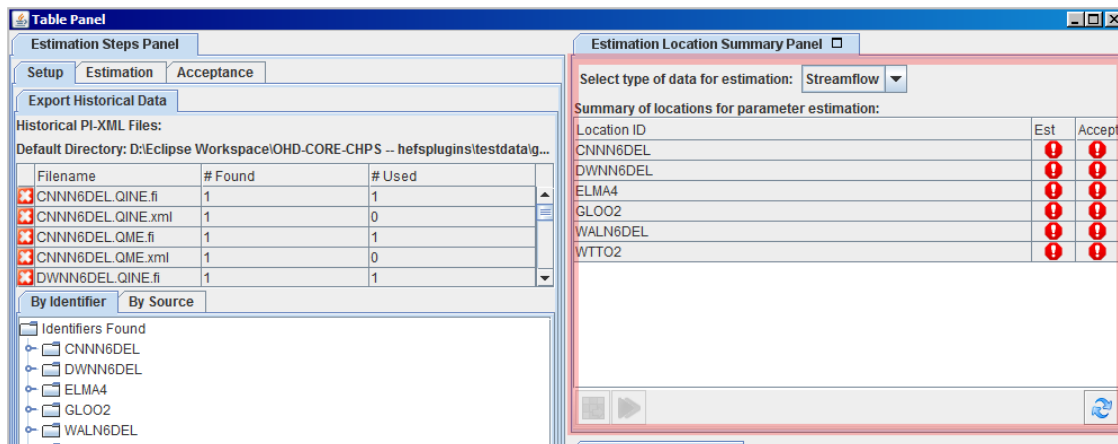
4.4.2 Usage

4.4.2.1 Performing a Parameter Estimation Step

A step is performed by making the corresponding step subpanel active, selecting EnsPost locations for which to perform the step, and clicking on the **Perform Step Button**.   Upon clicking **Perform Step**, a **Continue Dialog** will be displayed allowing the user to confirm or cancel the run; for example:
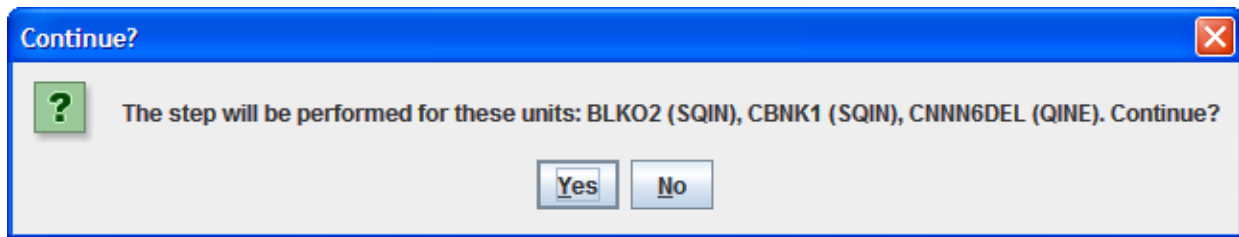
Click **Yes** to continue or **No** to cancel. If **Yes** is clicked, a **Step Progress Dialog** will be displayed providing the ability to cancel the step via a **Cancel Button**:
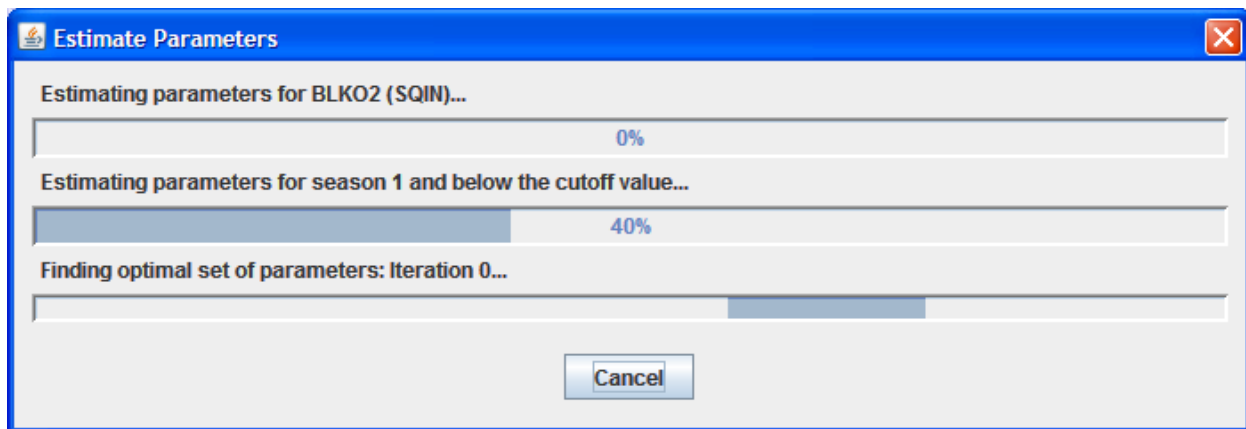
If the step fails for any reason, including if it was canceled, an error dialog will be displayed. If the step is only being performed for one EnsPost location or it is the last of multiple locations for which the step failed, then a **Step Failed Dialog** will be displayed explaining the cause of the failure:

Otherwise, an **Error Performing Step Dialog** will be displayed, giving the user the option to continue to the next selected EnsPost location (click **Yes** to continue, **No** to stop):

Error Performing Step!

⚠ Estimate Parameters step failed for BLKO2 (SQIN):
Error estimating HEFS EnsPost parameters: Unable to parse time series in file D:\workspace\HEFS-hefsplugins\testdata\guiEnsPostTestBed\piXMLFiles\BLKO2SQIN.xml. Message: File not found D:\workspace\HEFS-hefsplugins\testdata\guiEnsPostTestBed\piXMLFiles\BLKO2SQIN.xml (The system cannot find the file specified) at D:\workspace\HEFS-hefsplugins\testdata\guiEnsPostTestBed\piXMLFiles\BLKO2SQIN.xml
Continue to next location? (Click No to stop processing.)

[Yes]   [No]
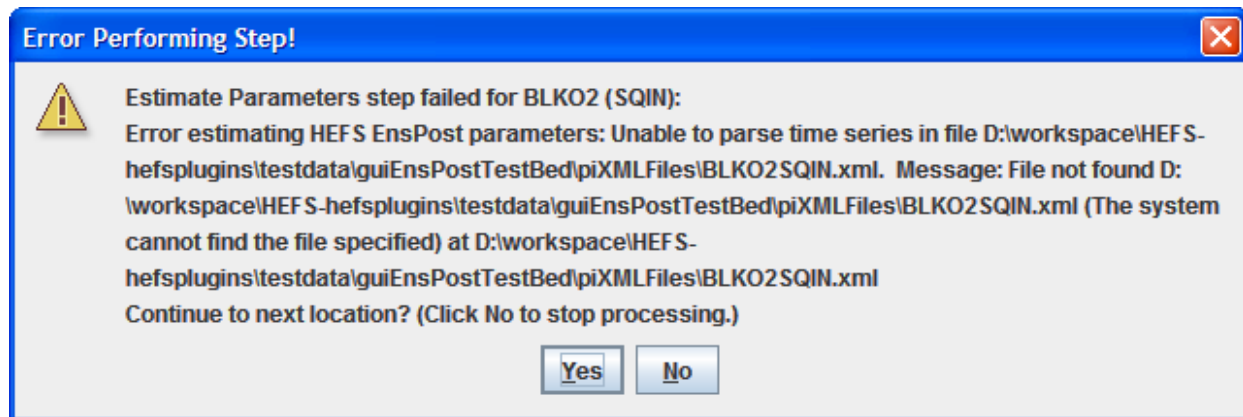
If the step is successful, the progress dialog will close with no additional dialog displayed. If a step is canceled by clicking on the **Cancel Button** in the **Step Progress Dialog**, the EnsPostPE may not immediately cancel the step. Rather, it will wait until the step can be canceled cleanly without causing any problems. Upon clicking **Cancel**, the button will be disabled until the step can be canceled.

4.5   Setup Subpanel

> **Tip**
>
> **The list of locations for which EnsPostPE can estimate parameters corresponds to the list of locationIds for which time series are found in the EnsPostPE run area.**

The **Setup Subpanel** of the **Estimation Steps Panel**, shown in Figure 11, allows the user to setup the EnsPost locations for which parameters are to be estimated. That is done by extracting time series from the localDataStore associated with the *parameter estimation SA* through the FEWS PI-service (see Section 4.2.4.3) and exporting those time series to PI-timeseries XML/FastInfoset (.fi) files within the following directory of the EnsPostPE run area:

<center><enspostpe_run_area>/piXMLFiles</center>

Alternatively, that directory can be populated manually following the instructions provided in the blue box in Section 4.2.1. Locations can be removed by removing corresponding files within that directory.

Subpanels within the **Setup Subpanel** are defined and are described below. The usage section explains how to perform each of these steps using the interface components provided.
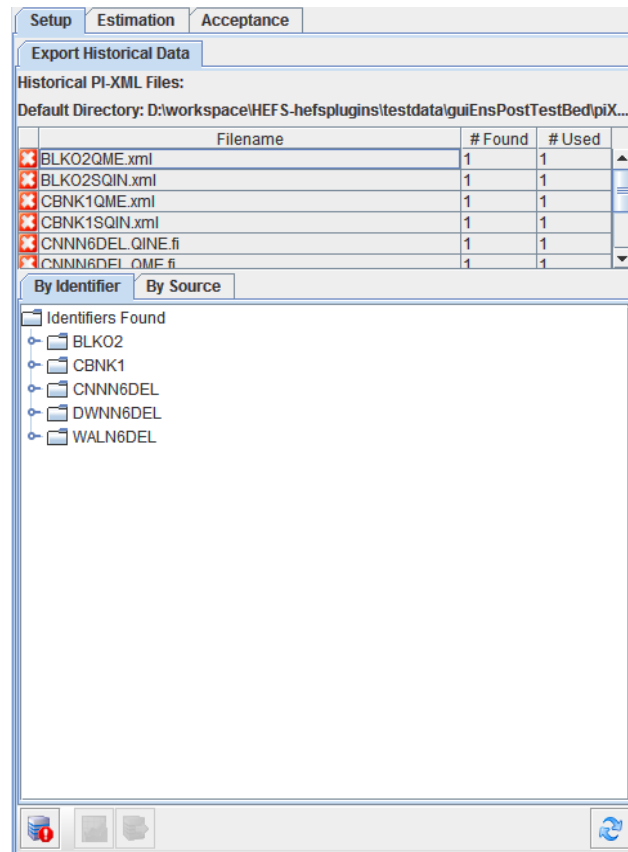
Figure 11: The **Setup Subpanel** of the **Estimation Steps Panel**.

4.5.1 Export Historical Data Subpanel

The **Export Historical Data Subpanel**, shown in Figure 11, facilitates creating, viewing, and quality controlling of the data. The time series are gathered by reading files from the directory

<p style="text-align: center;"><enspostpe_run_area>/piXMLFiles</p>

Available data is displayed in this components of the panel.

4.5.1.1 Components

- **Historical PI-XML Files Table**: Displays which files were found specifying historical simulated and observed streamflow time series. A delete column is included which, when clicked, removes the clicked row's file from the file system. A confirmation dialog will be shown before the file is deleted.
- **By Identifier Tree/By Source Tree**: The contents of the files found are listed in two trees displayed via a tabbed panel: the **By Identifier Tree** displays the time series first by EnsPost location identifier (locationId and parameterId XML elements of the PI-timeseries files), while the **By Source Tree** displays the time series first by source file found. The information provided in the tree includes locationId, parameterId, source file, start time, end time, and time step of the time series found. Both trees are selectable. Expand the tree nodes in order to view this information. For example:
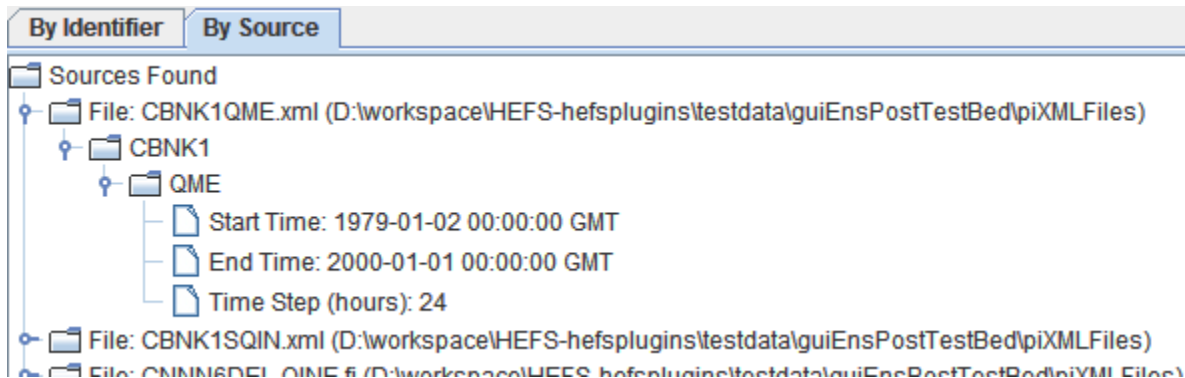
- **Reconnect to CHPS PI-service Button**: Click to open an **Enter Port Number Dialog** that allows for entering a port number to use for connecting to the PI-service:
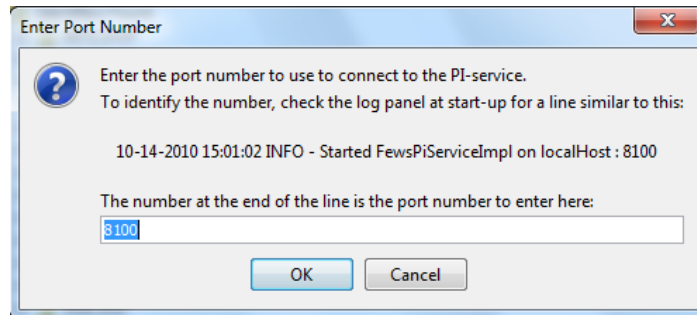
This is useful if the port number used previously (default is 8100) failed to yield a connection or connected to the wrong PI-service. A status icon is included within the larger icon, ✅ indicates a good connections, while ❗ indicates a bad connection. When the connection is bad, the **Export Time Series from CHPS DB Button** will be disabled. Detailed instructions for identifying this port number are provided in Section 4.2.4.3.

It is possible for the button to indicate a good connection even though it connected to the wrong PI-service. As such, follow the instructions for identifying the port number provided in Section 4.2.4.3 even when the icon indicates a good connection.

- **View Button**: Click to view time series selected from either the **By Identifier Tree** or **By Source Tree**. To view a time series, all selected nodes or leaves in the tree must be for the same CHPS locationId. See below for a description of the **Diagnostic Display Panel** that will be displayed.
- **Export Time Series from CHPS DB Button**: Click to extract time series from the localDataStore via the FEWS PI-service and export those time series to files on the file system for use by EnsPostPE. See Section 4.5.2.1 for more information.
- **Refresh Button**: Reread the in the EnsPostPE run area and reconstruct the trees. This needs to be clicked only if the files in the run area are modified manually while running EnsPostPE, as described in the blue box in Section 4.2.1.

4.5.1.2 Diagnostics

The diagnostics displayed in the **Diagnostic Display panel** for this subpanel are the time series as read in from the run area. The time series are displayed as dark and light blue lines in a Graphics Generator chart, which can display data in a table, allows for zooming, and includes an optional navigation panel at the bottom (see Section 4.9):
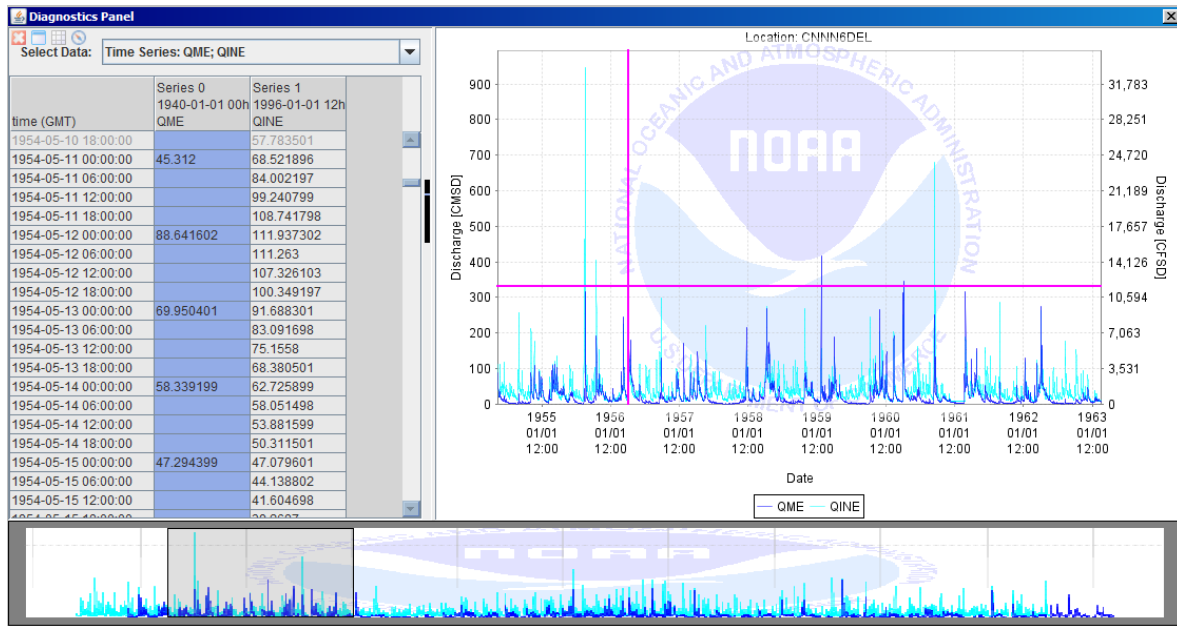
4.5.2 Usage

4.5.2.1 Exporting Historical Data via the FEWS PI-service

The standard mechanism for extracting historical simulated and observed time series from the localDataStore for use in estimating parameters is via the FEWS PI-service. To do so, do the following:

1. *Click on the* **Export Time Series from CHPS DB Button**.
Upon clicking, EnsPostPE will query the FEWS PI-service for a list of time series from which the user can make selections, displaying a progress dialog while the query is being performed. When the PI-service returns the needed information, the following **Select Time Series Dialog** will open:

The 'Export?' check box column for each row of the **Select Historical Time Series to Export Table** indicates if the time series will be acquired from the PI-service. The 'Done?' column indicates if data already exists in the EnsPostPE run area for that location.

2. *Select the time series to export.*
Check the checkbox in the 'Export?' column of the rows for each desired location and parameter. The buttons in the toolbar below the table can be used to check/uncheck all rows or check/uncheck all selected rows.

3. *Specify the date range of the data to acquire.*
Check on the **Start Date** and/or **End Date Checkboxes** and specify the start and/or end dates of the acquired time series in the date fields.

> - **Be sure to <u>exclude</u> data from the warm-up period of the historical simulation run when exporting simulated streamflow time series (see Section 4.2.1.1).**
> - If only one of the date range checkboxes is checked, then one end of the acquired time series will not be limited. For example, if the **Start Date Checkbox** is checked by not the **End Date Checkbox**, MEFPPE will acquire all data <u>after</u> the specified start date.

4. *Click **OK**.*
The EnsPostPE will then extract the time series and export them for use by EnsPostPE, displaying a progress dialog:
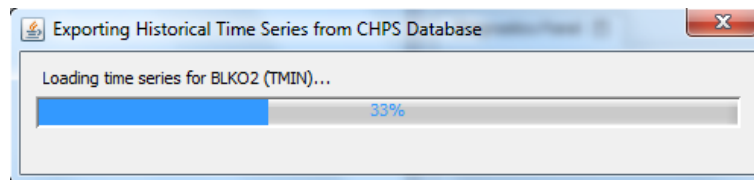
Upon completion, the progress dialog will close and the **By Identifier Tree/By Source Tree** within the **Historical Data SubPanel** will update to reflect the new files. Also, any new locations for which time series required for parameter estimation are now available will appear in the **Location Summary Panel** in the upper left corner of the **EnsPostPE Main Panel**.

4.5.2.2 Creating Historical Data Files Manually

See the blue box in Section 4.2.1 for complete instructions on preparing data for use in EnsPostPE manually and placing that data in the appropriate directory within the EnsPostPE run area.

4.6   Estimation Subpanel

The **Estimation Subpanel** of the **Estimation Steps Panel**, shown in Figure 12, is used to perform Step 2 of the parameter estimation procedure in Section 4.2.3: estimate parameters. The panel includes two subpanels: the **Locations Summary Subpanel** and **Estimation Options Subpanel**. Each subpanel and components are described below.

Figure 12 : The **Estimation Subpanel** of the **Estimation Steps Panel.**

4.6.1 Locations Summary Subpanel

The **Locations Summary Subpanel**, shown in Figure 12, summarizes the status of estimation for all locations and allows the user view log files, delete parameters, backup parameters, restore parameters, and select diagnostics to display.

4.6.1.1 Parameter File Backups

The EnsPostPE allows for *one* set of backup parameters per location. Whenever estimation is performed for selected EnsPost locations, if parameters have already been estimated for those locations, they will be backed-up, while any parameters that were backed-up will be discarded. Those newly backed-up parameters can later be restored if the new active parameters just estimated prove to be less desirable by clicking on the **Restore Parameters Button**.

4.6.1.2 Components

- **Summary of Estimated Parameters Availability Table**: Allows the user to select EnsPost locations for which to perform the step and view for which locations the step has been performed. It is a **Generic Summary Table** (Section 4.2.5.1) and includes all the standard buttons. Furthermore, it includes two additional columns: the 'Log File?' status column indicates if a log file is present for the current estimated parameters; the 'Backup?' status column indicates if backup parameters exist for the selected location.
- **View Log File Button**: Allows the user to view the contents of the log file for the estimated parameters. It opens up an **Estimation Log File Dialog** displaying the contents.
- **Load Parameters Button**: Click to load parameters for one selected location from the **Summary of Estimated Parameters Availability Table**. Upon loading, the **Parameters Summary Information Table** (below) will be updated to reflect the loaded parameters.
- **Restore Parameters Button**: Click to restore backup parameters for selected EnsPost locations. For the selected locations, the active and backed-up parameters will be swapped, making the backup parameters active and vice-versa. A **Continue? Dialog** will be opened allowing the user to confirm the restore.
- **Remove Parameters Button**: Click to remove the active parameters. A **Backup Parameters? Dialog** will open asking if the user wants to make the parameters backup parameters. If **Yes** is clicked, the parameters are backed-up. If **No** is clicked, the parameters are discarded. If **Cancel** is clicked, the remove is not performed.
- **Select Forecast Source Choice Box**: Allows the user to select the forecast source for which to view parameter summary information within the **Parameter Summary Information Table** (below).
- **Parameter Summary Information Table**: Displays the parameters loaded from the parameter files. The table allows for multiple selections. The following columns are provided:
    - 'Parameter Type': A descriptive name of the parameter.
    - '# Months': The number of months for which parameters were found.
    - 'Minimum': The smallest overall value found for the corresponding parameter.
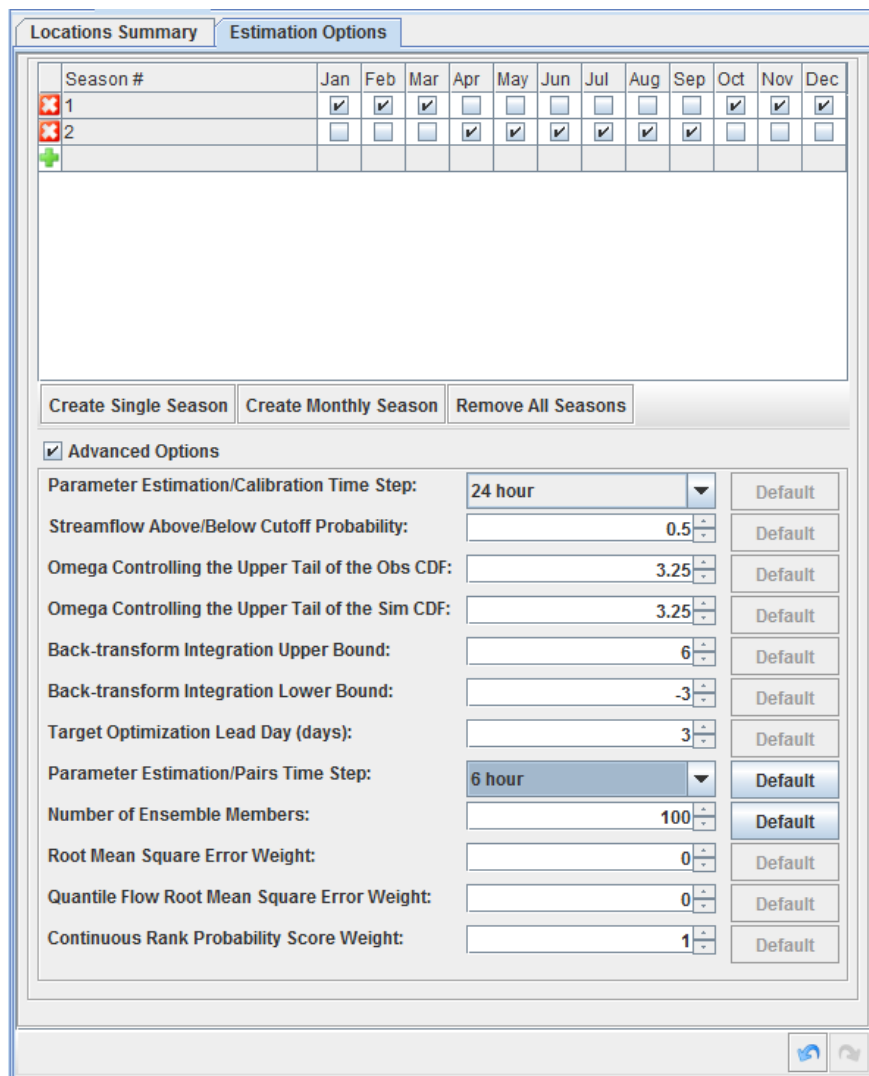   - 'Maximum': The largest overall value found for the corresponding parameter.

- **View Button**: Click to view the parameters for selected rows from the **Parameter Summary Information Table**. See the description of the diagnostics, below.
- **View CDF Button**: Click to view the CDFs by month for the EnsPost location. See the description of the diagnostics, below.

4.6.2 Estimation Options Subpanel

The **Estimation Options Subpanel** allows for users to specify options for the parameter estimation algorithm. All numerical parameters are edited using text fields with spinners; for example:

The number can be edited by clicking and typing or by clicking on the up or down arrows. The one choice-based option, **Parameter Estimation/Calibration Time Step**, is set by selecting the option from the choice box.
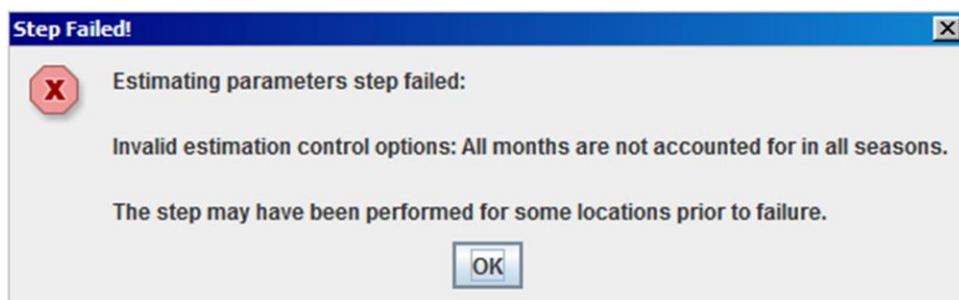
4.6.2.1 Seasons Table

The **Seassons Table** allows the user to specify multiple seasons for the estimation. The buttons below the table are described in Section 4.2.5.2. There are some additional buttons:

- Create Single Season Creates a single season with every month checked.
- Create Monthly Season Creates 12 seasons (1 for each month).
- Remove All Seasons Removes all seasons from the table.

In order for EnsPostPE to execute without errors, **every month must be in one and only one season**. Otherwise, parameter estimation will fail with this error:

Step Failed!

Estimating parameters step failed:

Invalid estimation control options: All months are not accounted for in all seasons.

The step may have been performed for some locations prior to failure.

OK

4.6.2.2 Advanced Options Subpanel

The EnsPostPE has several advanced options that are user defined via the **Advanced Options Panel** (see the screenshot provided above):

- **Parameter Estimation/Calibration Time Step**: The time step of parameter estimation, also called the *calibration time step*. Can be either 24-hours or 6-hours. (*Default value: 24-hours)*
- **Streamflow Above/Below Cutoff Probability:** Sets the probability associated with the streamflow cutoff value separating high flows from low flows. *(Default value: 0.5)*
- **Omega Controlling the Upper Tail of the Obs CDF:** Sets the omega parameter controlling the upper tail of the empirical observed cumulative distribution function beyond the largest observed value. *(Default value: 3.25)*
- **Omega Controlling the Upper Tail of the Sim CDF:** Sets the omega parameter controlling the upper tail of the empirical simulated cumulative distribution function beyond the largest simulated value. *(Default value: 3.25)*
- **Back Transform Integration Upper Bound:** Sets the upper bound on the region used to compute the numerical integration that is used within the back-transformation. *(Default value: 6.0)*
- **Back Transform Integration Lower Bound:** Sets the lower bound on the region used to compute the numerical integration that is used within the back-transformation. *(Default value: -3.0)*

To estimate parameters of the ERD and ERS error models, an optimization is performed; see Section 3.2. EnsPost is applied in order to generate hindcasts assuming the historical simulated time series provides a single-valued time series to post-process. The estimation options below control that optimization and the error measure that is minimized in order to identify the optimal parameters of the underlying ARX(1) model.

- **Target Optimization Lead Day (days):** Sets the number of lead days for which to apply EnsPost and compute the error message to be minimized as part of the optimization. *(Default value: 3)*
- **Parameter Estimation/Pairs Time Step:** If the *calibration time step* is 6-hours, then this controls the time step between successive hindcast T0s. For example, if the time step is set to "6 hour", then four hindcasts will be generated per day: one each with a T0 of 0Z, 6Z, 12Z, and 18Z. *(Default value: "24 hour")*
- **Number of Ensemble Members:** Sets the number of ensemble members generated per hindcast created during optimization. *(Default value: 2000)*
- **Root Mean Squared Error Weight:** Sets the weight to use for the computed root mean squared error when computing the error measure minimized during optimization. *(Default value: 0)*
- **Quantile Flow Root Mean Squared Error Weight:** Sets the weight to use for the quantile root mean square error when computing the error measure minimized during optimization. *(Default value: 0)*
- **Continuous Rank Probability Score Weight:** Sets the weight to use for the continuous rank probability score when computing the error measure minimized during optimization. *(Default value: 1)*

4.6.3 Diagnostics

The diagnostic display provided in the **Diagnostic Display Panel** when the **View Button** is clicked for the estimated parameters are designed primarily for parameter viewing and quality control. The diagnostic displays parameters selected within the **Parameter Summary Information Table**. The parameter values are displayed against the month of the year (1 – 12).

Provided below are two examples of diagnostic displays created when the ![icon] **View Button** is clicked. The first is for a single parameter (QINE Estimated Monthly Mean) while the second is for two parameters (for each additional parameter selected, an additional subplot is created in the **Diagnostic Display Panel**):
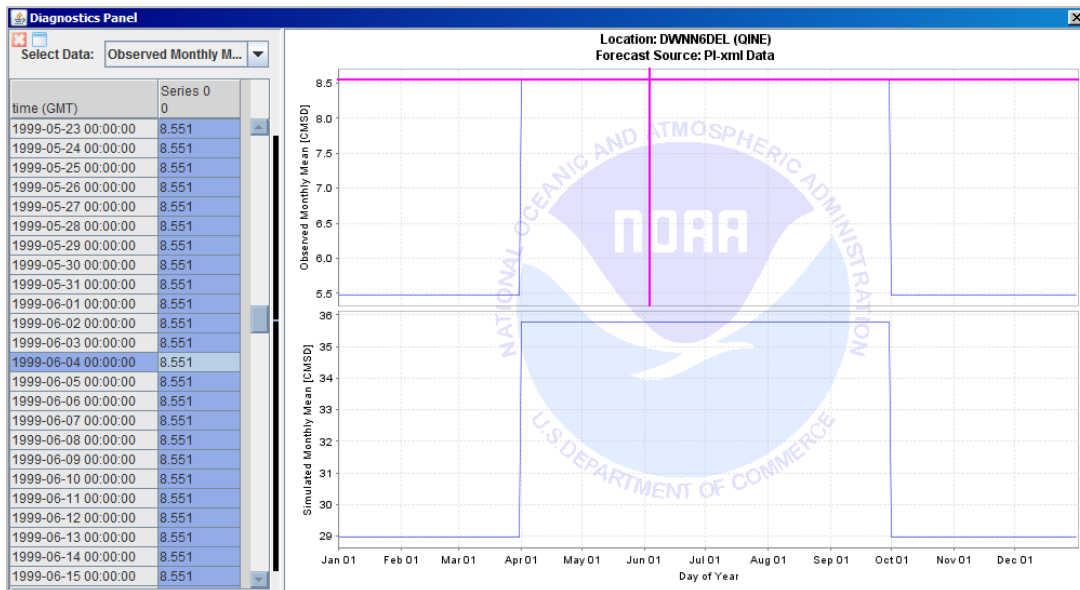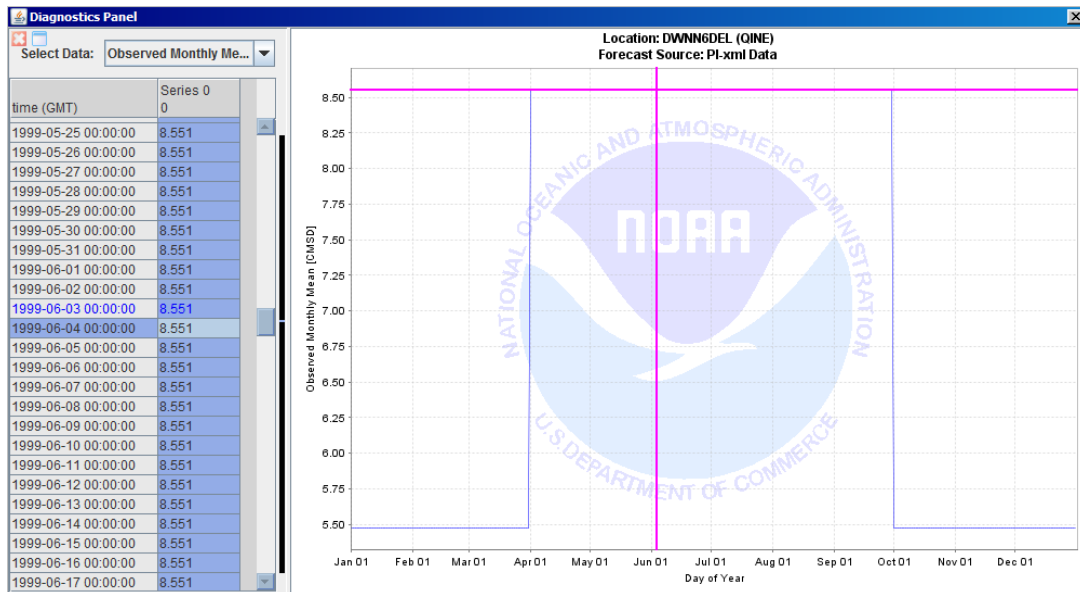
Provided below is an example of the diagnostic display created when the **View CDF Button** is clicked. It displays the empirical cumulative distribution functions estimated used for a forecast within a selected month, as estimated based on the seasons estimation options from the historical simulated streamflow time series and historical observed streamflow time series. Select the month for which to view the empirical cumulative distribution functions via the **Select Month of the Year Choicebox** at the top of the panel. CDFs viewed for months within the same season will be identical.
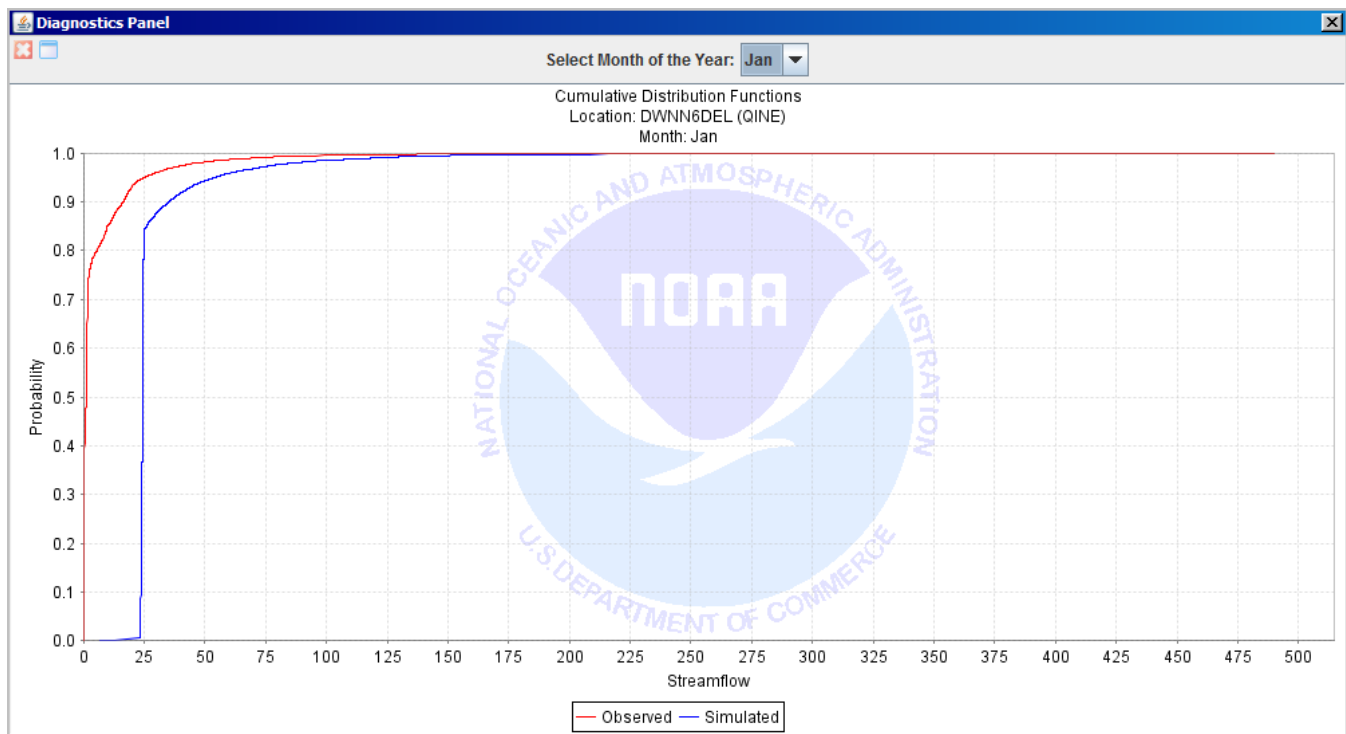
4.6.4 Usage

4.6.4.1 Estimating Parameters

To estimate parameters, select one or more EnsPost locations from the **Summary of Estimated Parameters Availability Table** and click on the **Perform Step Button**. See Section 4.4.2.1 for details on how to perform a step using the **Perform Step Button** of the **Estimation Steps Panel**.

Upon completion of estimating parameters, the following dialog will be displayed showing summary statistics of the parameter estimation for each selected location for which parameters were estimated:

LocationID	Season	Threshold	Months	Cutoff	Error Measure
CNNN6DEL	1	Below	Jan, Feb, Mar, Oct, Nov, Dec	24.012	2.03
CNNN6DEL	1	Above	Jan, Feb, Mar, Oct, Nov, Dec	24.012	8.214
CNNN6DEL	2	Below	Apr, May, Jun, Jul, Aug, Sep	34.046	3.397
CNNN6DEL	2	Above	Apr, May, Jun, Jul, Aug, Sep	34.046	8.685
DWNN6DEL	1	Below	Jan, Feb, Mar, Oct, Nov, Dec	24.389	1.713
DWNN6DEL	1	Above	Jan, Feb, Mar, Oct, Nov, Dec	24.389	2.565
DWNN6DEL	2	Below	Apr, May, Jun, Jul, Aug, Sep	24.534	1.162
DWNN6DEL	2	Above	Apr, May, Jun, Jul, Aug, Sep	24.534	4.599

OK

51

4.6.4.2 Loading Parameters and Viewing Diagnostics

To load parameters and view the diagnostics, do the following:

1. Select one EnsPost location (row) from the **Summary of Estimated Parameters Availability Table** of the **Locations Summary Subpanel**.
2. Click on the **Load Parameters Button**. A progress dialog will be displayed indicating that parameters are being loaded. This may take a minute or two. After completion, the **Parameter Summary Information Table** will update displaying the parameters loaded.
3. (Presently, there is only one source available, so this step can be skipped.) Select the forecast source for which you want to view parameters from the **Select Forecast Source Choice Box**. If the table is empty, then no parameters were loaded.
4. Select the estimated parameters (rows) that you wish to view from the **Parameter Summary Information Table**.
5. Click on the **View Button**.
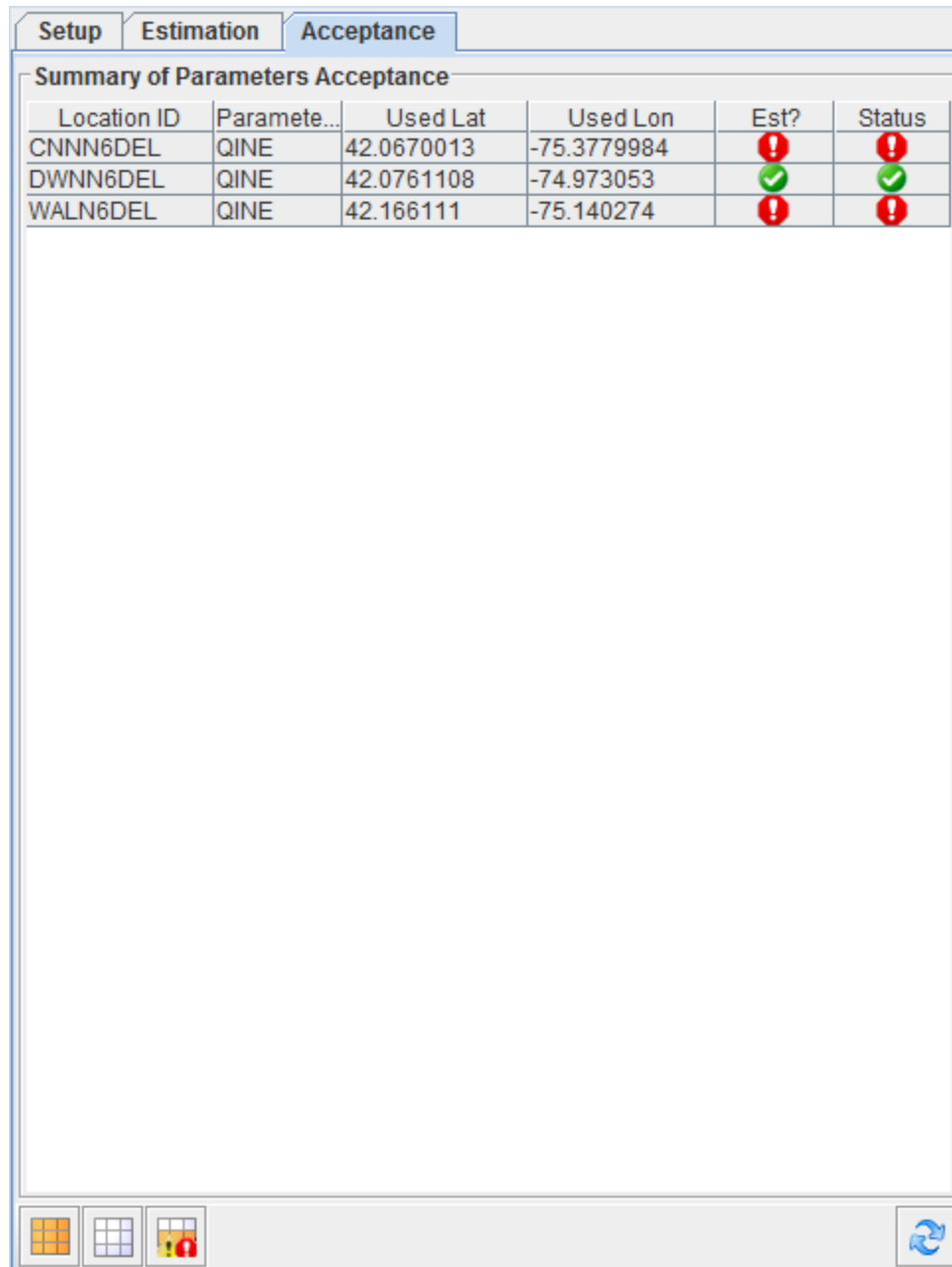
4.6.4.3 Backing-Up Parameters and Restoring Backup Parameters

To backup parameters, select one or more EnsPost locations from the **Summary of Estimated Parameters Availability Table** and click on the **Remove Parameters Button**. When the **Backup Parameters? Dialog** opens, click on **Yes**. Upon completion, the 'Status' column of the **Summary of Estimated Parameters Availability Table** will display a indicating that no active parameters are available, while the 'Backup?' column will display a indicating that backup parameters are available.

To restore backup parameters, select one or more EnsPost locations from the **Summary of Estimated Parameters Availability Table** and click on the **Restore Parameters Button**. When the **Continue? Dialog** opens, click on **Yes**. Upon completion, the 'Status' column of the **Summary of Estimated Parameters Availability Table** will display a or indicating that parameters are available, while the 'Backup?' column will display either a if there were active parameters when the button was clicked or if there were no active parameters.

4.7   Acceptance Subpanel

The **Acceptance Subpanel** of the **Estimation Steps Panel**, shown in Figure 13, is used to perform Step 3 of the parameter estimation procedure in Section 4.2.3: accept (tar and gzip) parameter files.

Figure 13 : The **Acceptance Subpanel** of the **Estimation Steps Panel**

4.7.1 About Parameter .tgz Files

As defined in Section 4.2.3, parameter .tgz files that are created will be placed in the directory

> *<ens_post_root_dir>*/ensPostParameters

and its file name will match the following pattern:

> \<locationID\>.\<parameterID\>.enspost.parameters.tgz

where locationId and parameterId are those XML elements associated with the selected location (as listed in the **Summary of Parameters Acceptance Table**.

4.7.2 Components

- **Summary of Parameters Acceptance Table**: Allows the user to select EnsPost locations for which accept parameters. It is a **Generic Summary Table** (Section 4.2.5.1) and includes all the standard buttons. In addition, the 'Est?' column is a status column and indicates if parameters are available to be accepted for the EnsPost location.

4.7.3 Usage

4.7.3.1 Accepting Parameters

To accept parameters for a particular location:

1. Select the row with the location ID in the Summary of Parameters Acceptance table.
2. Click on the **Perform Step Button**

See Section 4.4.2.1 for details on how to perform a step using the **Perform Step Button** of the **Estimation Steps Panel**.

4.8   Location Summary Panel

The **Location Summary Panel**, shown in Figure 14, summarizes the status of all steps to perform described in Section 4.2.3 for all EnsPost locations. The panel also includes a **Select Type of Data for Estimation Choice Box** for selecting the active estimation data type, a **Goto Step Panel Button** to facilitate quickly navigating the **Estimation Steps Panel**, and a **Run All Steps Button** to allow for performing all steps for multiple selected locations.
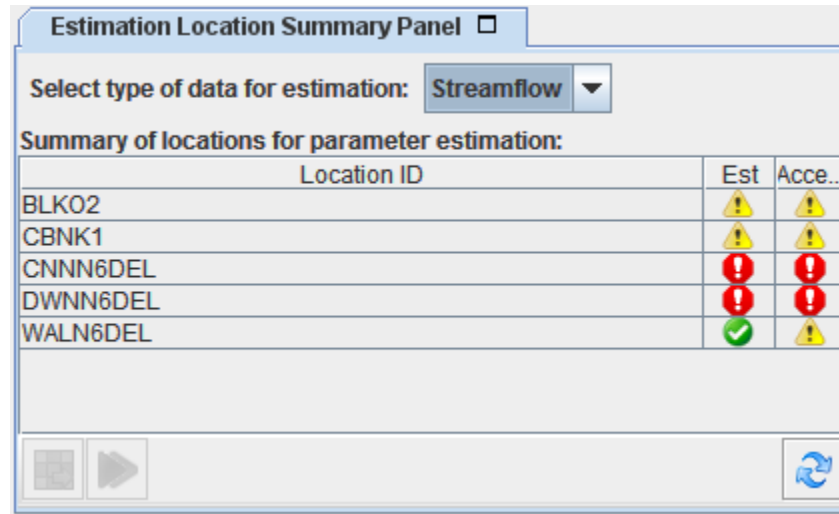
Figure 14 : The **Location Summary Panel**

4.8.1   Components

- **Select Type of Data for Estimation Choice Box**: Allows for selecting the active estimate data type ("Streamflow").
- **Summary of Locations for Parameter Estimation Table**: Summarizes the status of all steps to perform (except the setup step) for all EnsPost locations for the active estimation data type. The columns are as follows:
    - 'Location ID': The location id of the EnsPost location.
    - 'Est': Displays the status of Step 2: estimate parameters.
    - 'Accept': Displays the status of Step 3: accept (tar and gzip) parameter files.
- **Goto Step Panel Button**: Click to make the step subpanel corresponding to the selected column of the **Summary of Locations for Parameter Estimation Table** active within the **Estimation Steps Panel**. Also, if appropriate, it selects rows of the **Generic Summary Table** for that subpanel for the EnsPost locations selected within the **Summary of Locations for Parameter Estimation Table**. After clicking this button, the user should be able to click on the **Perform Step Button** to perform the step for all selected locations.
- **Run All Steps Button**: Click to run all of the steps for the selected EnsPost locations. Upon clicking a **Run All Options Dialog** will be displayed. Click on **OK** to continue with the run or **Cancel** to cancel it.
- **Refresh Button**: Click to refresh the status columns of the **Summary of Locations for Parameter Estimation Table**. This will also trigger a refresh of status columns in all subpanels of the **Estimation Steps Panel**.

4.8.2 Usage

4.8.2.1 Performing All Steps for Multiple EnsPost Locations

To perform all steps for desired EnsPost locations, do the following:

1. Select the rows for the desired EnsPost locations from the **Summary of Locations for Parameter Estimation Table**.
2. Click on the **Run All Steps Button**. A **Run All Options Dialog** will open:
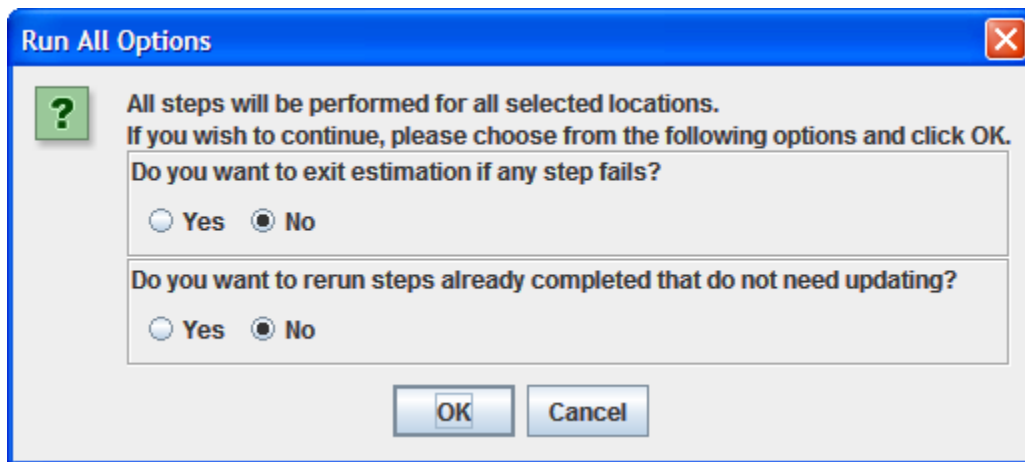
There are 2 options that the user can set:

o   Specify if the Run All job is to exit as soon as any step fails for any reason.
o   Specify if the Run All job should re-run already completed steps.

3. Set the options as desired and click **OK** (click **Cancel** to stop the run all). A progress dialog will open:
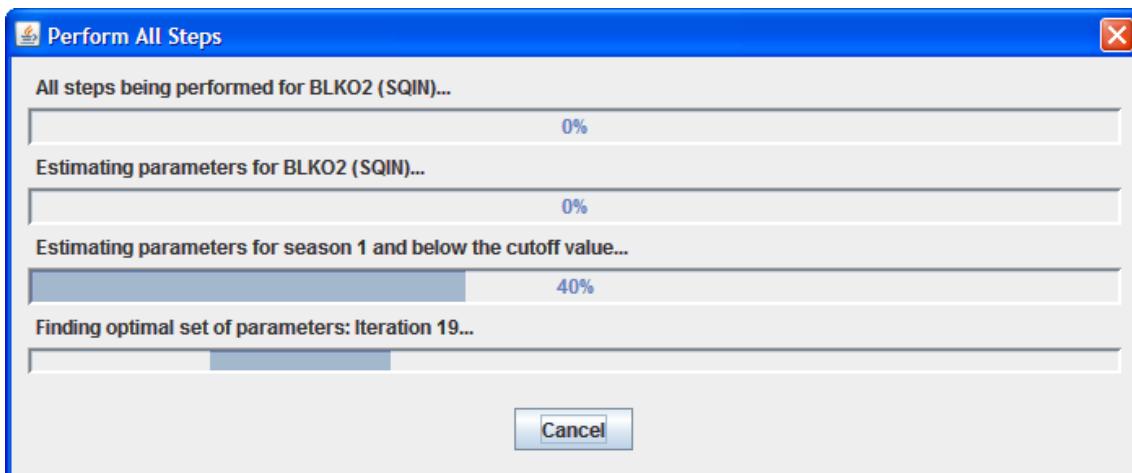
The run all can be canceled at any time by clicking on the **Cancel Button**. The run all will stop at the next opportunity and a **Perform All Errors Dialog** will be displayed where the last leaf will indicate that the user canceled the run:
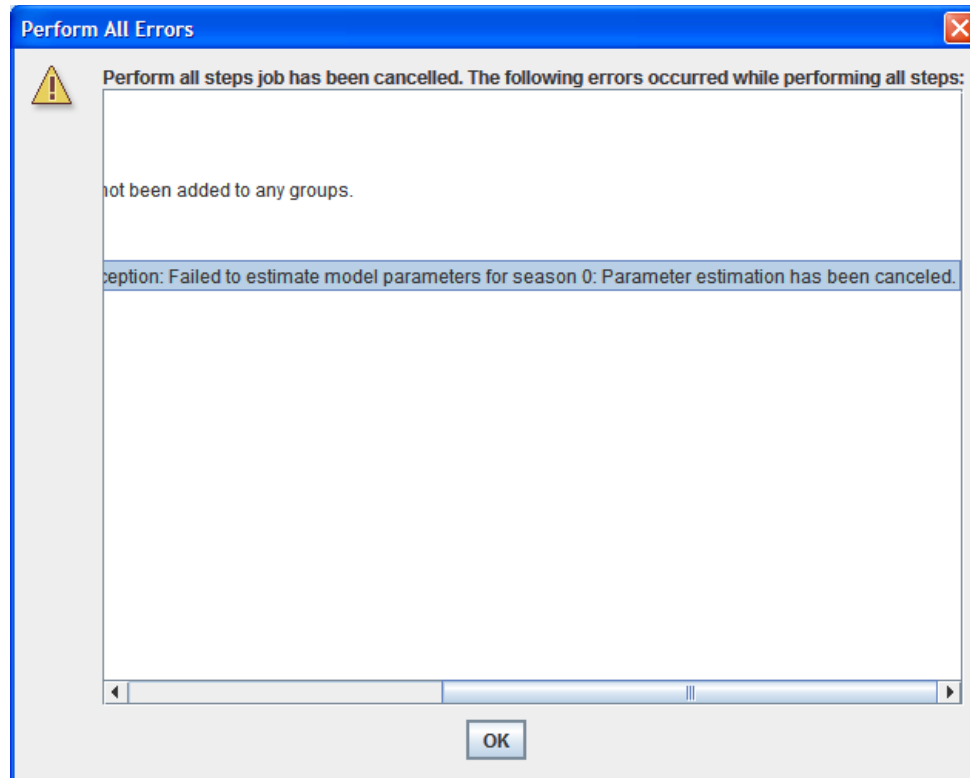
Click **OK** to close the dialog.

When a run all action is canceled, the steps performed prior to the cancelation will <u>not</u> be undone.

4. When the run all is completed, a **Perform All Errors Dialog** will be displayed indicating any errors that occurred during the run:
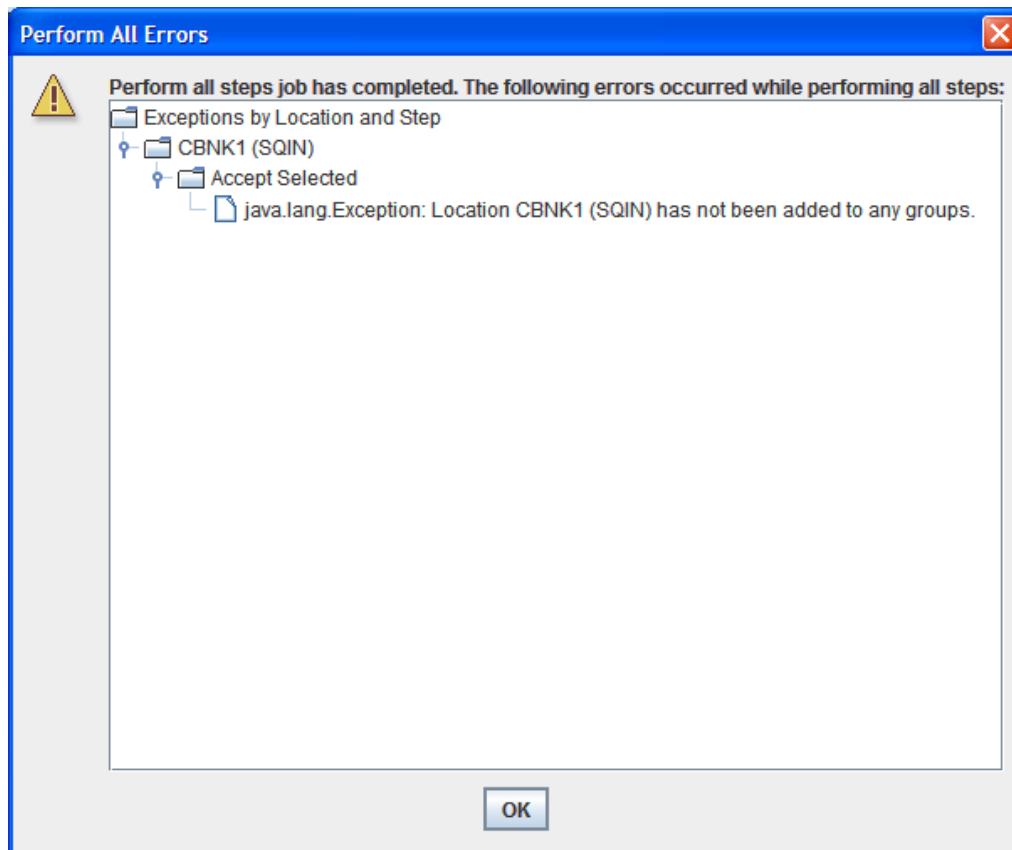
Expand the tree nodes in order to identify the errors that occurred. If no errors occurred, a message will be displayed indicating that no errors occurred.

5. Click **OK** to close the dialog.

4.9   Diagnostic Display Panel

The **Diagnostic Display Panel**, an example of which is shown in Figure 15, displays diagnostics as selected by the user via subpanels of the **Estimation Steps Panel**. The diagnostics that can be displayed are described with the subpanels.

All diagnostics in EnsPostPE are full featured Graphics Generator charts, which means that zooming is allowed, a table can be displayed by right clicking and clicking the **Show Table** menu item, and cross hairs will be drawn if a cell is clicked on in the table or a point is clicked on the chart (see Figure 15).

Some diagnostics can require a significant amount of time draw. Furthermore, it will need to be redrawn whenever the **Diagnostic Display Panel** is resized. To prevent a slowdown in the software resulting from spending too much time rendering displays, a ❌ **Dispose Button** is included in the upper left corner of the panel for all diagnostics to display. Click on the button to clear the panel.

The chart displayed in the **Diagnostic Display Panel** will not change until another diagnostic is selected to be displayed or the ❌ **Dispose Button** is clicked.
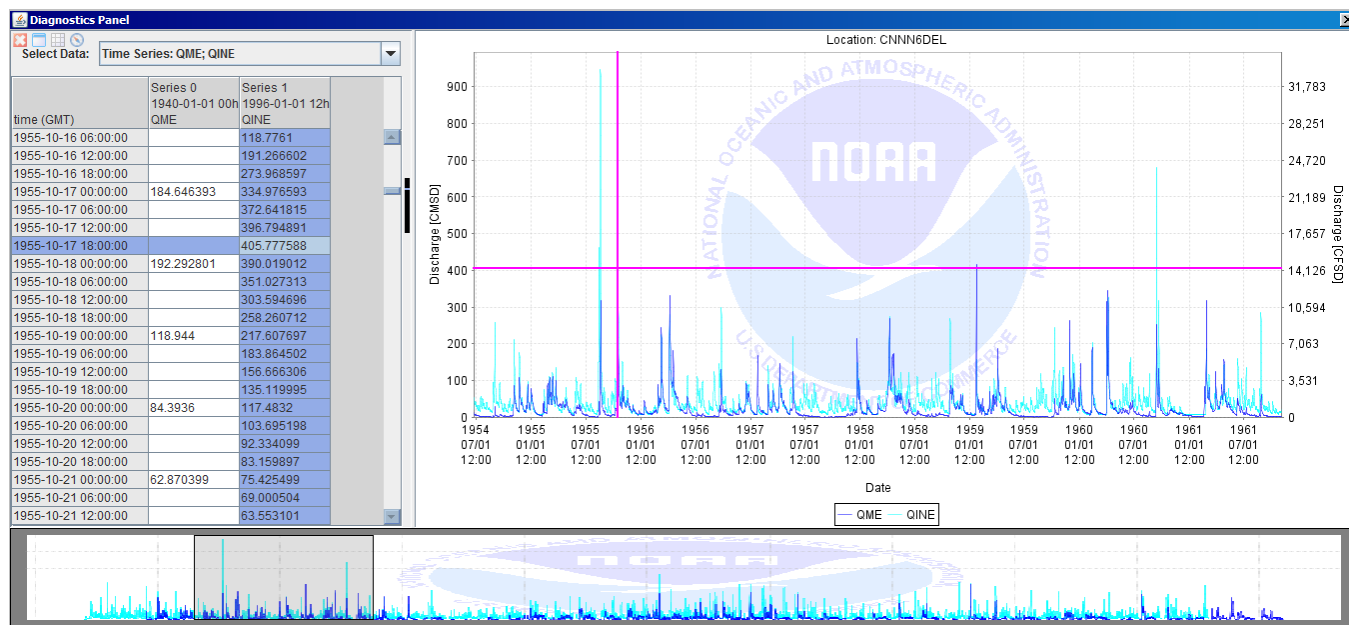
Figure 15 : The **Diagnostics Display Panel**

To avoid having to wait while a diagnostic is re-drawn, it is recommended that the **Diagnostic Display Panel** be sized prior to displaying any diagnostic. Furthermore, it is recommended that the panel be undocked to allow for it to be sized independently of the rest of EnsPostPE. As with any subpanel of the EnsPostPE, click on the square in the panel tab next to the name to do so:
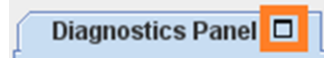
Diagnostics Panel ▢

4.9.1 Components

The only general components of a **Diagnostics Panel** are the buttons in the upper left corner, as shown in Figure 15. The buttons are only visible if a diagnostic is currently being displayed. The buttons are as follows:

- **Dispose Button**: Click on this button to dispose of the current diagnostics.
- **Move Diagnostic to Window Button**: Moves the diagnostic display to a separate window. This allows for multiple diagnostics to be displayed at the same time; after creating the first, click on this button to move the display to a window, allowing for a second to be drawn. This is not the same as undocking the **Diagnostics Panel**, which does not create a new panel, but, instead, detaches the existing panel.
- **Toggle Visibility of Data Table Button**: Shows or hides the **Data Table Panel**, described below. This is only visible in time series **Diagnostics Panels**.
- **Toggle Visibility of Navigation Panel Button**: Shows or hides the **Navigation Panel**, described below. This is only visible in time series **Diagnostics Panels**.

5   EnsPost Operational Reference Manual

5.1   HEFSEnsPostModelAdapter Reference Manual

5.1.1   Overview

This section of the manual describes how to use the HEFSEnsPostModelAdapter to operationally post-process ensemble streamflow forecasts in order to account for biases and hydrologic uncertainty. It is executed as a GeneralAdapter, requiring an appropriate module configuration file to be defined, with that module then being executed within an existing streamflow ensemble generating workflow. For instructions on configuring the adapter, see the *EnsPost Configuration Guide*.

The following input is required for the HEFSEnsPostModelAdapter to post-process a streamflow ensemble for a specific EnsPost location:

- a current (or most recent) observed streamflow value;
- an ensemble of forecast streamflow time series.

Based on the input combined with parameters calibrated prior to execution (See Section 4 for instructions how to run the parameter estimator), EnsPost generates a post-processed forecast ensemble, with each member in the input ensemble being used to generate a single member of the post-processed ensemble.

The HEFSEnsPostModelAdapter is capable of post-processing ensembles for multiple EnsPost locations through a single module configuration file.

5.1.1.1 Example Configuration

The following example configuration file is for execution of the HEFSEnsPostModelAdapter to post-process a forecast ensemble for the location CNNN6DEL(SQIN).

Example 1: Post-process a forecast ensemble for location CNNN6DEL(SQIN).

```
<generalAdapterRun xmlns="http://www.wldelft.nl/fews"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.wldelft.nl/fews http://fews.wldelft.nl/schemas/version1.0/generalAdapterRun.xsd">
    <general>
        <description>EnsPost CNNN6DEL</description>
        <piVersion>1.5</piVersion>
        <rootDir>%TEMP_DIR%</rootDir>
        <workDir>%ROOT_DIR%/work</workDir>
        <exportDir>%ROOT_DIR%/input</exportDir>
        <exportDataSetDir>%ROOT_DIR%/work/parameters</exportDataSetDir>
        <exportIdMap>IdExportENSPOST</exportIdMap>
        <importDir>%ROOT_DIR%/output</importDir>
        <importIdMap>IdImportENSPOST</importIdMap>
        <dumpFileDir>$GA_DUMPFILEDIR$</dumpFileDir>
```

```xml
    <dumpDir>%ROOT_DIR%</dumpDir>
    <diagnosticFile>%ROOT_DIR%/output/diag.xml</diagnosticFile>
</general>
<activities>

    <startUpActivities>
        <makeDir>
            <dir>%ROOT_DIR%/work</dir>
        </makeDir>
    </startUpActivities>

    <exportActivities>

        <exportTimeSeriesActivity>
            <exportFile>inputs.xml</exportFile>
            <timeSeriesSets>
                <timeSeriesSet> <!-- THE OBSERVED TIME SERIES -->
                    <moduleInstanceId>ADDSUB_CNNN6DEL_ADDSPILL_Forecast</moduleInstanceId>
                    <valueType>scalar</valueType>
                    <parameterId>QIN</parameterId>
                    <locationId>CNNN6TOT</locationId>
                    <timeSeriesType>simulated forecasting</timeSeriesType>
                    <timeStep unit="hour" multiplier="6" />
                    <relativeViewPeriod unit="hour" start="-120"
                        startOverrulable="true" end="0" />
                    <readWriteMode>read only</readWriteMode>
                </timeSeriesSet>
                <timeSeriesSet> <!-- THE ENSEMBLE TO POST-PROCESS -->
                    <moduleInstanceId>ADDSUB_CNNN6DEL_ADD.USQ_Forecast</moduleInstanceId>
                    <valueType>scalar</valueType>
                    <parameterId>SQIN</parameterId>
                    <locationId>CNNN6TIF</locationId>
                    <timeSeriesType>simulated forecasting</timeSeriesType>
                    <timeStep unit="hour" multiplier="6" />
                    <relativeViewPeriod unit="hour" start="0" end="120"
                        endOverrulable="true" />
                    <readWriteMode>add originals</readWriteMode>
                    <ensembleId>MEFP</ensembleId>
                    <ensembleMemberIndexRange start="1961"
                        end="1997" />
                </timeSeriesSet>
            </timeSeriesSets>
        </exportTimeSeriesActivity>

        <exportRunFileActivity>
            <exportFile>%ROOT_DIR%/run_info.xml</exportFile>
            <properties>
                <int key="printDebugInfo" value="0" />
                <string key="errorModel" value="ERS" />
                <string key="parameterDir" value="$ENS_POST_ROOT_DIR$/ensPostParameters"/>
                <string key="parameterFile" value="CNNN6.SQIN.enspost.parameters.tgz"/>
            </properties>
        </exportRunFileActivity>

    </exportActivities>
```

```
<executeActivities>
   <executeActivity>
      <command>
         <className>ohd.hseb.hefs.enspost.adapter.HEFSEnsPostModelAdapter</className>
         <binDir>$OHDBINDIR$</binDir>
      </command>
      <arguments>
         <argument>%ROOT_DIR%/run_info.xml</argument>
      </arguments>
      <timeOut>60000</timeOut>
   </executeActivity>
</executeActivities>

<importActivities>
   <importTimeSeriesActivity>
      <importFile>outputs.xml</importFile>
      <timeSeriesSets>
         <timeSeriesSet>
            <moduleInstanceId>CNNN6DEL_ENSPOST_Forecast</moduleInstanceId>
            <valueType>scalar</valueType>
            <parameterId>SQIN</parameterId>
            <locationId>CNNN6DEL</locationId>
            <timeSeriesType>simulated forecasting</timeSeriesType>
            <timeStep unit="hour" multiplier="6" />
            <readWriteMode>add originals</readWriteMode>
            <ensembleId>HEFSENSPOST</ensembleId>
            <ensembleMemberIndexRange start="1961"
               end="1997" />
         </timeSeriesSet>
      </timeSeriesSets>
   </importTimeSeriesActivity>
</importActivities>
   </activities>
</generalAdapterRun>
```

5.1.2 Model Parameters

The HEFSEnsPostModelAdapter does not use GeneralAdapter model parameters. Rather, the parameters are estimated through the EnsPostPE and the adapter is pointed to them through run file porperties.

5.1.3 Model Run File Properties

Some model run file properties can be defined by EnsPost location, overriding default properties but only for that location. In such cases, the location for which the property applies is specified as a suffix after a '.' within the name of the property; for example:

```
<int key="parameterFileName.CNNN6DEL" value="CNNN6DEL.SQIN.parameters.tgz"/>
```

In this example, the name of the EnsPost .tgz parameter file to use for the location CNNN6DEL (an MARFC example) is set to CNNN6DEL.SQIN.parameters.tgz. The name of each location specific property is denoted in the table below with the suffix "[.<locationId>]".

The following exportRunFileActivity properties can be defined (for the Default Value column, "(required)" indicates that the property is required, so that there is no default value):

- For all properties of type "Boolean", if the value of the property is "true", then the property is true. Any other value is treated as false.

Property Name	Type	Default Value	Description / Example
disaggOutput	Yes/No	Yes	If No, then the post-processed streamflow ensemble will be output at the *calibration time step* of the EnsPost parameters. That time step is specified through the estimation options of the EnsPostPE. If Yes, then the post-processed streamflow ensemble will be disaggregated, or downscaled, to the time step of the streamflow ensemble provided exported to the adapter. See Section 3.3.
errorModel	String	(required)	Specify the ErrorModel for EnsPost to use. Valid options are "ER0", "ERD", and "ERS". See Section 3.4.
parameterDir	String	(required)	The directory where .tgz parameter files (see the parameterFileName property below) will be found. Typically, this should be set to the following: $ENS_POST_ROOT_DIR$/ensPostParameters
parameterFile[.<locationId>]	String	See Description	Allows the parameter file name to be manually specified by the user. The default parameter file name is for an EnsPost location is the following: <locationId>.<parameterId>.enspost.parameters.tgz where the locationId and parameterId are those provided in the streamflow ensemble to post-process. **If the locationId is not included as a suffix for the property, then the specified parameter file will be used for <u>all</u> ensembles to post-process.**

multiLocationRun	Yes/No	No	This must be set to Yes if EnsPost is to execute for multiple locations. The appropriate input time series must then be provided for each EnsPost location for which to execute, which likely requires the use of moduleInstanceSet and locationSet when defining the exportTimeSeriesActivity. Furthermore, the importTimeSeriesActivity will also likely require the use of sets.
testing	String (the value must a long/integer, but it cannot be an int property)	(not used)	Used for testing/debugging: when present, the random number generator used in the ERS model is seeded to the value of this property. This causes the random number generator to always generate the same sequence of numbers. Use of this property allows for reproducing operational results. When not set, the seed used is determined based on the computer system time. That seed is output to the log (diag.xml) in a line that appears similar to this following example: "Random number generator initialized with the seed 1407855596551" To reproduce the results of a run, then, add the "testing" string property to the module configuration file with a value equal to the specified number; using the example above: <string key="testing" value="1407855596551">
useStratifiedRandomSampling	true/ false	false	If false, then simple random sampling will be used in the ERS model in order to sample from the error distribution. Otherwise, stratified random sampling will be used across ensemble members to ensure a more even spread across the entire distribution of errors.
peLocationID (deprecated; do not use)	String	No	**This property is no longer supported.**
peParameterID (deprecated; do not use)	String	No	**This property is no longer supported.**
printDebugInfo	Integer	0	If positive, debug level log messages will be output by the adapter to the diagnostics (diag.xml) file for ingest by CHPS.

5.1.4 Model Input Time Series

The exportTimeSeriesActivity XML element in the configuration file exports time series that indicate for which EnsPost locations (locationId) ensembles will be post-processed and provides the require time series to determine the latest observed value and the ensemble to post-process.

For each EnsPost location for which an ensemble is to be post-processed, the following time series must be exported via the exportTimeSeriesActivity XML element:

- **Observed streamflow time series (one):** A time series of observed stream flow values, typically QIN or QME, ending on or after the forecast T0.
- **Forecast ensemble streamflow time series (many):** An ensemble (one time series per member) of forecast streamflow values, typically SQIN, QINE, or SQME, starting either on T0 or one time step after T0.

> Executing EnsPost for multiple locations (segments) through a single configuration file is only allowed if the run file property multiLocationRun is set to 'Yes'.

For each time series, the time step of the time series <u>must not</u> be larger than the *calibration time step* of the parameters to be applied. If the time step is less than the *calibration time step*, then the EnsPost will aggregate the time series to the *calibration time step* using a mean calibration. **If the data type after aggregation (if necessary) is not compatible with the data type of parameter estimation, then an error will occur.** See Section 3.3 for an explanation of data compatibility and Table 1 for examples illustrating data compatibility.

For example, if parameter estimation was performed at a 6-hour *calibration time step* using 6-hour QIN (historical observed) and 6-hour SQIN (historical simulated) time series, then the resulting parameters must be applied to *instantaneous* 6-hour data. However, if the HEFSEnsPostModelAdapter is executed using 1-hour QIN for the observed streamflow time series, then that will be aggregated by the model to be 6-hour *mean* observed streamflow, and the parameters will not be applicably because it expects 6-hour *instantaneous* data.

The following is an example of a valid exportTimeSeriesActivity for one location for which parameters were estimated with a 24-hour *calibratin time step*:

```
<exportTimeSeriesActivity>
  <exportFile>inputs.xml</exportFile>
  <timeSeriesSets>
    <timeSeriesSet>
      <moduleInstanceId>ADDSUB_CNNN6DEL_ADDSPILL_Forecast</moduleInstanceId>
      <valueType>scalar</valueType>
      <parameterId>QIN</parameterId>
      <locationId>CNNN6TOT</locationId>
      <timeSeriesType>simulated forecasting</timeSeriesType>
      <timeStep unit="hour" multiplier="6" />
      <relativeViewPeriod unit="hour" start="-120"
```

```
                startOverrulable="true" end="0" />
          <readWriteMode>read only</readWriteMode>
        </timeSeriesSet>
        <timeSeriesSet>
          <moduleInstanceId>ADDSUB_CNNN6DEL_ADD.USQ_Forecast</moduleInstanceId>
          <valueType>scalar</valueType>
          <parameterId>SQIN</parameterId>
          <locationId>CNNN6TIF</locationId>
          <timeSeriesType>simulated forecasting</timeSeriesType>
          <timeStep unit="hour" multiplier="6" />
          <relativeViewPeriod unit="hour" start="0" end="120"
            endOverrulable="true" />
          <readWriteMode>add originals</readWriteMode>
          <ensembleId>MEFP</ensembleId>
          <ensembleMemberIndexRange start="1961"
            end="1997" />
        </timeSeriesSet>
      </timeSeriesSets>
    </exportTimeSeriesActivity>
```

5.1.5 Model Execution

The following executeActivity XML element should always be included within the executeActivities XML element in the configuration file (see the example in Section 5.1.1.1):

```
    <executeActivity>
      <command>
        <className>ohd.hseb.hefs.enspost.adapter.HEFSEnsPostModelAdapter</className>
      </command>
      <arguments>
        <argument>%ROOT_DIR%/run_info.xml</argument>
      </arguments>
      <timeOut>300000</timeOut>
    </executeActivity>
```

Only the timeOut XML element, set to 5 minutes (300,000 milliseconds) above, should be modified if it is determined that it is not sufficient for the configured adapter run.

5.1.6 Model Output Time Series

The ensemble streamflow time series output by the model adapter will have the same header information (locationId, parameterId, start time, end time, etc.) as the ensemble streamflow forecast time series provided as input, except that it will be given the ensemble id of "HEFSENSPOST". Further, if the disaggOutput property is set to 'No', then the time step of the *output* ensemble will match the calibration time step of the parameters applied. Otherwise, it will match the time step of the *input* ensemble.

The following is an example of an importTimeSeriesActivity XML element for a single location for which the output was disaggregated to the time step of the input ensemble:

```
    <importActivities>
      <importTimeSeriesActivity>
        <importFile>outputs.xml</importFile>
```

```
<timeSeriesSets>
  <timeSeriesSet>
    <moduleInstanceId>CNNN6DEL_ENSPOST_Forecast</moduleInstanceId>
    <valueType>scalar</valueType>
    <parameterId>SQIN</parameterId>
    <locationId>CNNN6DEL</locationId>
    <timeSeriesType>simulated forecasting</timeSeriesType>
    <timeStep unit="hour" multiplier="6" />
    <readWriteMode>add originals</readWriteMode>
    <ensembleId>HEFSENSPOST</ensembleId>
    <ensembleMemberIndexRange start="1961"
        end="1997" />
  </timeSeriesSet>
</timeSeriesSets>
</importTimeSeriesActivity>
</importActivities>
```

> **notes**
>
> Do not change the importFile XML element's value in the importTimeSeriesActivity.
> It should always match that shown in the example above.

5.1.7 Model Description

The HEFSEnsPostModelAdapter algorithm follows this general process in order to gather time series for locations and call the EnsPost model:

1. Parse the run file properties.
2. Parse the provided time series in order to identify the EnsPost locations. If mulitiple locations are identified but the multiLocationRun property is not set to 'Yes', then error out.
3. For each location…
   a. Load the parameters for that locationId and parameterId. See the description of the parameterDir and parameterFile run file properties in Section 5.1.3.
   b. Aggregate time series to the *calibration time step* if necessary. If the resulting data types (accumulative or instantaneous data) are not consistent with the parameters, error out.
   c. Execute the user specified error model for the given location. The error model is specified via the errorModel run file property described in Section 5.1.3.
   d. Disaggregate the time series to the time step of the input ensemble if the disaggOutput run file property is set to 'Yes'. See Section 5.1.3.
   e. Record the generated ensemble in a results list.
4. Output the results list to the outputs.xml file so that CHPS can import it as instructed in the module configuration file.

> **notes**
>
> The model adapter is fully compatible with FEWS id-mapping capabilities. The parameter file searched for will be determined based on the external locationId and parameterId defined in the id-mapping.

6   REFERENCES

Brown, J. D., and D.-J. Seo, 2012: Evaluation of a nonparametric post-processor for bias correction and uncertainty estimation of hydrologic predictions. Hydrological Processes, doi:10.1002/hyp.9263

Bogner K, and F. Pappenberger 2011: Multiscale error analysis, correction, and predictive uncertainty estimation in a flood forecasting system, Water Resources Research, W07524, DOI: 10.1029/2010WR009137.

Chen, S.-T., and P.-S. Yu, 2007: Real-time probabilistic forecasting of flood stages, Journal of Hydrology, 340 (1–2), 63–77.

Hantush, M.M., and L. Kalin, 2008: Stochastic residual-error analysis for estimating hydrologic model predictive uncertainty, Journal of Hydrologic Engineering, 13(7):585-596.

Krzysztofowicz, R., 1999: Bayesian Theory of Probabilistic Forecasting Via Deterministic Hydrologic Model, Water Resources Research, 35(9), 2739-2750.

Montanari, A., and A. Brath, 2004: A stochastic approach for assessing the uncertainty of rainfall-runoff simulations, Water Resources Research, 40, W01106, doi:10.1029/2003WR002540.

Montanari, A., and G. Grossi, 2008: Estimating the uncertainty of hydrological forecasts: A statistical approach, Water Resources Research, 44, W00B08, doi:10.1029/2008WR006897.

Regonda et al., 2012: Short-term Ensemble Streamflow Forecasting Using Operationally-Produced Single-valued Streamflow Forecasts - A Hydrologic Model Output Statistics (HMOS) Approach (in preparation)

Seo, D.-J., H. Herr, and J. Schaake, 2006: A statistical post-processor for accounting of hydrologic uncertainty in short-range ensemble streamflow prediction, Hydrological Earth System Science Discussions, 3, 1987-2035.

Todini, E., 2008: A model conditional processor to assess predictive uncertainty in flood forecasting, International Journal of River Basin Management, 6(2), 123–137.